

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: June 16, 2004, 19:06:57 ; Search time 84.1687 Seconds
(without alignments)
2265.920 Million cell updates/sec

le: US-09-332-063-2

fect score: 3347

quence: 1 MPRAPSSASQFVADPPA.....KTPPIQLQEPDAMVEYLI 675

ring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

al number of hits satisfying chosen parameters: 1586107

imum DB seq length: 0
imum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : 1: Genesecp1980s.*
2: Genesecp1990s.*
3: Genesecp2000s.*
4: Genesecp2001s.*
5: Genesecp2002s.*
6: Genesecp2003as.*
7: Genesecp2003bs.*
8: Genesecp2004s.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB	ID	Description
1	3347	100.0	675	3	AA54052	Aay54052 An angiot
2	3331	99.5	993	7	AAE37918	Aae37918 Human CGD
3	3325	99.3	675	3	AA54053	Aay54053 A variant
4	2381	71.1	479	5	ABP43965	Abp43965 Unidentif
5	1448	43.3	637	5	ABP69085	Abp69085 Human pol
6	1447	43.2	608	4	AAU69439	AAu69439 Human EUR
7	1441	43.1	801	7	ADC38518	Adc38518 Human AML
8	1441	43.1	869	7	ADC37654	Adc37654 Human ang
9	1400	41.8	882	7	ABR56262	AbR56262 Mouse Jun
10	1090.5	32.6	772	7	ABR56263	AbR56263 Mouse Jun
11	1040.5	31.1	849	3	AAAB41861	AAb41861 Human ORF
12	1037	31.0	467	5	ABG97502	Abg97502 Human NOV
13	987.5	29.5	706	4	ABA92866	ABa92866 Human pro
14	903	27.0	759	4	ABG17959	Abg17959 Novel hum
15	765.5	22.9	562	7	ADC37657	Adc37657 Human ang
16	712	21.3	340	4	AAAM3770	AAm3770 Human pol
17	663	19.8	143	3	AA54054	Aay54054 Angiostat
18	521.5	15.6	277	6	ABU70751	ABu70751 Human adi
19	328.5	9.8	182	4	AAAB94793	AAb94793 Human pro
20	308	9.2	1072	5	ABBS7023	ABb57023 Mouse iac
21	302	9.0	1444	4	ABG15667	Abg15667 Novel hum
22	296.5	8.9	1013	4	ABBY1039	ABb71039 Drosophil
23	286	8.5	1071	4	ABBB60579	ABb60579 Drosophil
24	285.5	8.5	566	4	ABBB1040	ABb61040 Drosophil
25	284.5	8.5	2346	4	ABBB63519	ABb63519 Drosophil

ALIGNMENTS

RESULT 1

AA54052

ID AAY54052 standard; protein; 675 AA.

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

XX AC AAY54052;

be used to manufacture medicaments for treating angiogenesis-related diseases or disorders, such as tumor conditions, diabetes, rheumatoid arthritis, and even some inflammatory diseases such as psoriasis, chronic inflammation of the intestine, asthma, etc. The protein may also be able to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene therapy techniques

Sequence 675 AA;

Query Match 100.0%; Score 3347; DB 3; Length 675;
 Best Local Similarity 100.0%; Pred. No. 1.2e-201;
 Matches 675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MPRAQPSASQYQVPADPPFAIVSRAQQMVEILSDENRNLROELGECYKVARLQKVETI 60
 1 MPRAQPSASQYQVPADPPFAIVSRAQQMVEILSDENRNLROELGECYKVARLQKVETI 60

61 QRYSEAYENLVKSSSKREALEKAMRNKLEGEIRRMDFNDRLETRLETANKOLAEKEYEG 120
 61 QRYSEAYENLVKSSSKREALEKAMRNKLEGEIRRMDFNDRLETRLETANKOLAEKEYEG 120

121 SEDTRKTIISQFAKNKESQREKEKLEAEATASSTNEDQRRHIEIRDOALSNAQAKVKVL 180
 121 SEDTRKTIISQFAKNKESQREKEKLEAEATASSTNEDQRRHIEIRDOALSNAQAKVKVL 180

181 EELKKQVYVVKVKKMOQALVQLQACREKLEHRLTRLEERLESRIQORQNCOP 240
 181 EELKKQVYVVKVKKMOQALVQLQACREKLEHRLTRLEERLESRIQORQNCOP 240

241 TNVSEYNAALMELLREKEERILALEADMTKQKYLEENVVHFDALDAATVAQRODT 300
 241 TNVSEYNAALMELLREKEERILALEADMTKQKYLEENVVHFDALDAATVAQRODT 300

301 VISHSPNTSYDTALEARIQKEEIEILMANKRCIDMEGRITKLHAQIIKXDMIKVLOQRS 360
 301 VISHSPNTSYDTALEARIQKEEIEILMANKRCIDMEGRITKLHAQIIKXDMIKVLOQRS 360

361 RKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKDDKSWKSGILIG 420
 361 RKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKDDKSWKSGILIG 420

421 GDYRAEVVPTSPVPSTPILLSAHSKTSRDCSTQTERGTSNKTAAVAPISVPAPVA 480
 421 GDYRAEVVPTSPVPSTPILLSAHSKTSRDCSTQTERGTSNKTAAVAPISVPAPVA 480

481 AATAAATATAATITTTWAAAAPVAVAAAAAATAAATAAATAAATAAATAAATAAATAA 540
 481 AATAAATATAATITTTWAAAAPVAVAAAAAATAAATAAATAAATAAATAAATAAATAA 540

541 AASVASAAAVAPSAAAAAAQQVAPAPAPVAPALVVPAPAPALVVPAPAPALVVPAPAP 600
 541 AASVASAAAVAPSAAAAAAQQVAPAPAPVAPALVVPAPAPALVVPAPAPALVVPAPAP 600

601 VAPTEPPTTPPAQAQAVPASPATGPGPHRLSPSLTCNPKDGTGVPVHNTLERTPIQ 660
 601 VAPTEPPTTPPAQAQAVPASPATGPGPHRLSPSLTCNPKDGTGVPVHNTLERTPIQ 660

661 ILGQEPDAEMVEYLI 675
 661 ILGQEPDAEMVEYLI 675

SULT 2
 E37918
 AAE37918 standard; protein; 993 AA.

AAE37918;

06-NOV-2003 (first entry)

Human CGDD-7 protein.

Human; cell growth, differentiation and death protein; CGDD; leukaemia;

neurodegenerative disorder; Parkinson's disease; Alzheimer's disease; muscular disorder; myotonic dystrophy; catatonia; endocrine disorder; diabetes; Grave's disease; cancer; immunological disorder; scleroderma; systemic lupus erythematosus; allergy; Crohn's disease; renal disorder; gastrointestinal disorder; Goodpasture's syndrome; infection; cirrhosis; cardiovascular disorder; atherosclerosis; hepatic disease; transgenic; transgenic animal; gene therapy; neuroprotective; relaxant; cytostatic; dermatological; immunosuppressive; cerebroprotective; anticonvulsant; antibacterial; antiparasitic; fungicide; virucide; uropathic; cardiant; protozoacide; nootropic.

Homo sapiens.

W02003050253-A2.
 19-JUN-2003.
 04-DEC-2002; 2002MO-US039133.
 07-DEC-2001; 2001US-0340747P.
 20-DEC-2001; 2001US-0342761P.
 15-JAN-2002; 2002US-0349705P.
 06-FEB-2002; 2002US-0354764P.
 12-FEB-2002; 2002US-0356216P.

(INCY-) INCYTE GENOMICS INC.

Griffin JA, Rankumar J, Emerling BM, Kable AZ, Elliott VS, Marquis JP, Baughn ME, Gorvad AE, Yue H, Lee EA, Becha SP, Tang YT, Tran UK, Swarnakar A, Lee S, Ison CH, Hafalia AJA, Tran B, Sprague WW, Lee SY, Khare R, Gandhi AS, Gietzen KJ, Bhatia U, Burdill JD, Blake JJ, Ho A, Zheng W;

WPI: 2003-532903/50.
 N-PSDB; AAD57227.

New CGDD polypeptides, useful for diagnosing, preventing, and treating disorders associated with an abnormal expression or activity of CGDD, e.g. neuromuscular, immunological, cardiovascular disorders, cancer and/or infections.

Claim 1; Page 220-222; 299pp; English.

The present invention relates to novel cell growth, differentiation and death (CGDD) proteins and polynucleotides encoding them. The sequences of the invention are useful in diagnosing, preventing and treating disorders associated with an abnormal expression or activity of CGDD such as neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's disease), muscular disorders (e.g. myotonic dystrophy, catatonia), endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g. leukaemia, cervical or breast cancers), immunological disorders (e.g. scleroderma, systemic lupus erythematosus, allergies), gastrointestinal disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's syndrome), infections (e.g. viral, bacterial, fungal, parasitic, protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis) and hepatic diseases (e.g. cirrhosis). The polynucleotides can be used to create humanised animals or transgenic animals to model human diseases. The invention is also used in gene therapy. The present sequence is human CGDD-7 protein

Sequence 993 AA;

Query Match 99.5%; Score 3331; DB 7; Length 993;
 Best Local Similarity 99.8%; Pred. No. 2e-200; 2; Indels 0; Gaps 0;
 Matches 672; Conservative 1; Mismatches 1;

Oy 1 MPRAQPSASQYQVPADPPFAIVSRAQQMVEILSDENRNLROELGECYKVARLQKVETI 60
 Db 319 MPRAQPSASQYQVPADPPFAIVSRAQQMVEILSDENRNLROELGECYKVARLQKVETI 378
 Oy 61 QRYSEAYENLVKSSSKREALEKAMRNKLEGEIRRMDFNDRLETRLETANKOLAEKEYEG 120
 Db 379 QRYSEAYENLVKSSSKREALEKAMRNKLEGEIRRMDFNDRLETRLETANKOLAEKEYEG 438

121 SEDTRKTIISQLPAKFKESQREKEKLEAELATARSTNEDORRHIEIRDOALSNAQAKVVKL 180
121 SEDTRKTIISQLPAKFKESQREKEKLEAELATARSTNEDORRHIEIRDOALSNAQAKVVKL 180
439 SEDTRKTIISQLPAKFKESQREKEKLEAELATARSTNEDORRHIEIRDOALSNAQAKVVKL 498
439 SEDTRKTIISQLPAKFKESQREKEKLEAELATARSTNEDORRHIEIRDOALSNAQAKVVKL 498
181 EELKKKKQVYVDKVKKKQALVOLQAACEKREQLERLRLTRERLESRLIQORQNCQP 240
181 EELKKKKQVYVDKVKKKQALVOLQAACEKREQLERLRLTRERLESRLIQORQNCQP 240
499 EELKKKKQVYVDKVKKKQALVOLQAACEKREQLERLRLTRERLESRLIQORQNCQP 558
499 EELKKKKQVYVDKVKKKQALVOLQAACEKREQLERLRLTRERLESRLIQORQNCQP 558
241 TNVSEYNAALMELLREKEERILALEADMTKWKQKYLEENVMRHPALDAAATVAARQDIT 300
241 TNVSEYNAALMELLREKEERILALEADMTKWKQKYLEENVMRHPALDAAATVAARQDIT 300
559 TNVSEYNAALMELLREKEERILALEADMTKWKQKYLEENVMRHPALDAAATVAARQDIT 618
559 TNVSEYNAALMELLREKEERILALEADMTKWKQKYLEENVMRHPALDAAATVAARQDIT 618
301 VISHSPNTSYDTALEARIQKEEELIWMANKRCIDMEGRITLHAQIIKDMIKVLQORS 360
301 VISHSPNTSYDTALEARIQKEEELIWMANKRCIDMEGRITLHAQIIKDMIKVLQORS 360
619 VISHSPNTSYDTALEARIQKEEELIWMANKRCIDMEGRITLHAQIIKDMIKVLQORS 678
619 VISHSPNTSYDTALEARIQKEEELIWMANKRCIDMEGRITLHAQIIKDMIKVLQORS 678
361 RKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDKSKWKSGLIGL 420
361 RKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDKSKWKSGLIGL 420
679 RKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDKSKWKSGLIGL 738
679 RKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDKSKWKSGLIGL 738
421 GDYRAEYVPSTPSPVPSTPILLSAHSKSGSDCSTQTERGTSNKTAAPISVPAPVAA 480
421 GDYRAEYVPSTPSPVPSTPILLSAHSKSGSDCSTQTERGTSNKTAAPISVPAPVAA 480
739 GDYRAEYVPSTPSPVPSTPILLSAHSKSGSDCSTQTERGTSNKTAAPISVPAPVAA 798
739 GDYRAEYVPSTPSPVPSTPILLSAHSKSGSDCSTQTERGTSNKTAAPISVPAPVAA 798
481 AATAAAITATAATITTTMVAAPVAVAAAAAPAAAAAPSPATAAAATAAVSPAAGQIPA 540
481 AATAAAITATAATITTTMVAAPVAVAAAAAPAAAAAPSPATAAAATAAVSPAAGQIPA 540
799 AATAAAITATAATITTTMVAAPVAVAAAAAPAAAAAPSPATAAAATAAVSPAAGQIPA 858
799 AATAAAITATAATITTTMVAAPVAVAAAAAPAAAAAPSPATAAAATAAVSPAAGQIPA 858
541 AASVASAAAAPGSAAGAAAVQVAPAPAPVPAPALVPVPAPAAQASAPAQOAPTSA 600
541 AASVASAAAAPGSAAGAAAVQVAPAPAPVPAPALVPVPAPAAQASAPAQOAPTSA 600
859 AASVASAAAAPGSAAGAAAVQVAPAPAPVPAPALVPVPAPAAQASAPAQOAPTSA 918
859 AASVASAAAAPGSAAGAAAVQVAPAPAPVPAPALVPVPAPAAQASAPAQOAPTSA 918
601 VAPTPATPTPAVAQAEVPAAPATGPGPHRLSPSLTICNPDKTDGVFHSNTLERTPIQ 660
601 VAPTPATPTPAVAQAEVPAAPATGPGPHRLSPSLTICNPDKTDGVFHSNTLERTPIQ 660
919 VAPTPATPTPAVAQAEVPAAPATGPGPHRLSPSLTICNPDKTDGVFHSNTLERTPIQ 978
919 VAPTPATPTPAVAQAEVPAAPATGPGPHRLSPSLTICNPDKTDGVFHSNTLERTPIQ 978
661 ILGQEPDAEMVYVLI 675
661 ILGQEPDAEMVYVLI 675
979 ILGQEPDAEMVYVLI 993
979 ILGQEPDAEMVYVLI 993

SULT 3
Y34053
AAY54053 standard; protein; 675 AA.
AAY54053;
27-MAR-2000 (first entry)
A variant of an angiogenesis-associated protein which binds plasminogen.
Human; angiogenesis-associated protein; plasminogen; ABP-1;
kringle domain; angiotensin; plasminogen receptor;
angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
obesity; gene therapy.
Homo sapiens.
Key Location/Qualifiers
Misc-difference 135
/label= An, Ser, Asp
Misc-difference 148..150
/note= "these residues are either Glu-Leu-Ala or Thr-Thr-Pro"
W0996038-A1.
23-DEC-1999.
11-JUN-1999; 99WO-EP004109.

15-JUN-1998; 98SE-00002130.
15-JUN-1998; 98US-0089266P.
17-DEC-1998; 98SE-00004372.
29-DEC-1998; 98US-0114386P.
XX (PEAA) PHARMACIA & UPJOHN AB.
XX Holmgren L, Troyanovsky B;
XX WPI; 2000-106099/09.
DR Novel human protein useful for treating angiogenesis associated diseases
PT or disorders.
PT Claim 5; Page 45-49, 58pp; English.
XX The present sequence represents a polymorphic variant of a human
XX angiogenesis-associated protein which is able to bind an N-terminal
XX fragment of plasminogen. The native protein, described in AAY54052, is
XX designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or
XX kringle 5 (K5) of plasminogen. These four kringle domains comprise
XX angiotensin. The ABP-1 protein acts as a receptor for plasminogen. The
XX angiotensin-binding domain of the ABP-1 protein is described in AAY54054.
XX ABP-1 can be used to manufacture medicaments for treating angiogenesis-
XX related diseases or disorders, such as tumor conditions, diabetes,
XX rheumatoid arthritis, and even some inflammatory diseases such as
XX psoriasis, chronic inflammation of the intestine, asthma, etc. . The
XX protein may also be able to treat and cure, or prevent, obesity. The ABP-
XX 1 DNA can be used in gene therapy techniques
XX Sequence 675 AA;
Query Match 99.3%; Score 3325; DB 3; Length 675;
Best Local Similarity 99.4%; Pred. No. 2.9e-200;
Matches 671; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 MPRAQPSASQVPADPPFAIVSRAQOMVEILSDENRNLROELGCVKVARLQKQVETEI 60
Db 1 MPRAQPSASQVPADPPFAIVSRAQOMVEILSDENRNLROELGCVKVARLQKQVETEI 60
Qy 61 QRVSEAYENLVKSSKREALEKAMRNKLEGEIRRMDFNEDLRLERLETANKQLAKEYEG 120
Db 61 QRVSEAYENLVKSSKREALEKAMRNKLEGEIRRMDFNEDLRLERLETANKQLAKEYEG 120
Qy 121 SEDTRKTIISQLPAKFKESQREKEKLEAELATARSTNEDORRHIEIRDOALSNAQAKVVKL 180
Db 121 SEDTRKTIISQLPAKFKESQREKEKLEAELATARSTNEDORRHIEIRDOALSNAQAKVVKL 180
Qy 181 EELKKKKQVYVDKVKKKQALVOLQAACEKREQLERLRLTRERLESRLIQORQNCQP 240
Db 181 EELKKKKQVYVDKVKKKQALVOLQAACEKREQLERLRLTRERLESRLIQORQNCQP 240
Qy 241 TNVSEYNAALMELLREKEERILALEADMTKWKQKYLEENVMRHPALDAAATVAARQDIT 300
Db 241 TNVSEYNAALMELLREKEERILALEADMTKWKQKYLEENVMRHPALDAAATVAARQDIT 300
Qy 301 VISHSPNTSYDTALEARIQKEEELIWMANKRCIDMEGRITLHAQIIKDMIKVLQORS 360
Db 301 VISHSPNTSYDTALEARIQKEEELIWMANKRCIDMEGRITLHAQIIKDMIKVLQORS 360
Qy 361 RKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDKSKWKSGLIGL 420
Db 361 RKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDKSKWKSGLIGL 420
Qy 421 GDYRAEYVPSTPSPVPSTPILLSAHSKSGSDCSTQTERGTSNKTAAPISVPAPVAA 480
Db 421 GDYRAEYVPSTPSPVPSTPILLSAHSKSGSDCSTQTERGTSNKTAAPISVPAPVAA 480
Qy 481 AATAAAITATAATITTTMVAAPVAVAAAAAPAAAAAPSPATAAAATAAVSPAAGQIPA 540
Db 481 AATAAAITATAATITTTMVAAPVAVAAAAAPAAAAAPSPATAAAATAAVSPAAGQIPA 540
Qy 541 AASVASAAAAPGSAAGAAAVQVAPAPAPVPAPALVPVPAPAAQASAPAQOAPTSA 600

541 AASVASAAVAPSAASAAVQVAPAPAPVPAPALVPVPAFAAAQASAPAOQAPTSAPA 600
601 VAPTPAPTPPAVAQAQEVPAAPGAPGPHRLSLPSLTCPDKTDGVPVHSNLTLEKTPIQ 660
602 VAPTPAPTPPAVAQAQEVPAAPGAPGPHRLSLPSLTCPDKTDGVPVHSNLTLEKTPIQ 660
661 ILGQEPDAEMVEYLI 675
661 ILGQEPDAEMVEYLI 675

SULT 4
P43965
ABP43965 standard; protein; 479 AA.

ABP43965;

26-FEB-2003 (first entry)

Unidentified protein sequence.

Neuroprotective; immunomodulator; cancer; chromosome X; cytostatic;
anti-inflammatory; gene therapy; nutritional supplement; wound; burn;
ulcer; Alzheimer's disease; Huntington's disease;
amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
vulnary.

Homo sapiens.

WO200231111-A2.

18-APR-2002.

11-OCT-2001; 2001WO-US027760.

12-OCT-2000; 2000US-00687527.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
Xue AJ, Yang Y, Wehrman T, Drmanac R;

WPI; 2002-426278/45.

N-PSDB; ABQ61209.

New polypeptides and their encoded proteins, useful as nutritional
sources or supplements, or in gene therapy, particularly for treating
wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
inflammation.

Claim 20; SEQ ID # 868; 357pp + Sequence Listing; English.

The invention relates to 446 newly isolated polynucleotide sequences. The
activity of polynucleotides of the invention may be described as,
vulnary, neuroprotective, immunomodulator, cytostatic and anti-
inflammatory. Compositions comprising nucleic acids of the invention are
useful for treating a mammalian subject, or as nutritional sources or
supplements. These are useful in gene therapy, particularly for treating
wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
amyotrophic lateral sclerosis, autoimmune disorders, cancer or
inflammation. The nucleic acids and polypeptides are also useful in
diagnostic and research methods. The sequences given in records ABP43544-
ABP43989 represent polypeptides encoded by polynucleotides of the
invention. NOTE: The sequence data for this patent did not form part of
the printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pcr_sequences

Sequence 479 AA;

Query Match 71.1%; Score 2381; DB 5; Length 479;
Best Local Similarity 100.0%; Pred. No. 3.4e-141;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 MQQALVOLQAACEKREQLBHLRLTRLESLRLESLRQRCNQCTNVSYNAAALMELLR 256
DB 1 MQQALVOLQAACEKREQLBHLRLTRLESLRLESLRQRCNQCTNVSYNAAALMELLR 60
QY 257 EKBERILALEADMTKWEQKYLEENVVHFAALDAATVAAQRODTTVISHSPNTSYDTALEA 316
DB 61 EKBERILALEADMTKWEQKYLEENVVHFAALDAATVAAQRODTTVISHSPNTSYDTALEA 120
QY 317 RIQKEEEIILMANKRCLDMGRIKTLHAQIIIEKDAIKVYLQQRSRKEPSKTBOLSCMRPA 376
DB 121 RIQKEEEIILMANKRCLDMGRIKTLHAQIIIEKDAIKVYLQQRSRKEPSKTBOLSCMRPA 180
QY 377 KSLMSISNAGSGLLSHSSTLTGSPIMEEEDDKSWKSGSLGILLGCDYRAEYVPSPTSPVP 436
DB 181 KSLMSISNAGSGLLSHSSTLTGSPIMEEEDDKSWKSGSLGILLGCDYRAEYVPSPTSPVP 240
QY 437 PSTPLLSAHSKTSGRDCSTQTERGTSNKTAAVAPISVPAPVAAATAATAATATATIT 496
DB 241 PSTPLLSAHSKTSGRDCSTQTERGTSNKTAAVAPISVPAPVAAATAATAATATATIT 300
QY 497 TWVAAAPVAAAAAASPAATAATAATAATAATAATAATAATAATAATAATAATAATA 556
DB 301 TWVAAAPVAAAAAASPAATAATAATAATAATAATAATAATAATAATAATAATAATA 360
QY 557 AAQVQVAPAAAPVAPVAPALVPVPAFAAAQASAPAOQAPTSAPAVAPTPTTAPAQA 616
DB 361 AAQVQVAPAAAPVAPVAPALVPVPAFAAAQASAPAOQAPTSAPAVAPTPTTAPAQA 420
QY 617 EYPASPATGPHRLSLPSLTCPDKTDGVPVHSNLTLEKTPIQILGQEPDAEMVEYLI 675
DB 421 EYPASPATGPHRLSLPSLTCPDKTDGVPVHSNLTLEKTPIQILGQEPDAEMVEYLI 479

RESULT 5

ABP69085

ID ABP69085 standard; protein; 637 AA.

AC ABP69085;

DT 20-JAN-2003 (first entry)

DE Human polypeptide SEQ ID NO 1132.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
cell-proliferative disorder; neurodegenerative disease; bacterial;
Parkinson's disease; Alzheimer's disease; autoimmune disease;
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
arthritis; cytostatic; immunomodulator; neurotropic; dermatological;
antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;
antiarthritic.

OS Homo sapiens.

PN WO200270539-A2.

XX 12-SEP-2002.

PD 05-MAR-2002; 2002WO-US005095.

PF 05-MAR-2001; 2001US-00799451.

PR (HYSE-) HYSEQ INC.

XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki Y, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac R;

XX WPI; 2002-759812/82.

DR N-PSDB; ABZ11302.

PT New polynucleotides comprising sequences assembled from expressed

122 EDTRKTSIQLEFAKNKESOREKELEAEALATARNSTNEDORRHIEIRDOALSNAQAKVVKLE 181
 130 ED-KAAEGHYASQNKFEKLEKELEMELEAAVRTASEDHRRHIELDOALSNAQAKVVKLE 248
 182 BELKKQYVVDKVMQOALVOLQAAACEKEQLEHRLRTRLEHRLSLRTOQCGNCQPT 241
 249 BELREKQAYVEKVKLQALQALQALQALQALQALQALQALQALQALQALQALQALQALQ 308
 242 NVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVAARDTTV 301
 309 NMPEYNAPALLELVREKEERILALEADMTKWEQKYLEESTIRHFANNAATAAARDTTI 368
 302 ISHSPNTSY-DTALREARIQKEEBEILMANRCLDMGRITKLHAQIIKEDAMIKVLOORS 360
 369 INHNRNGSYGSSLEAHIMWEEBEEVVOANRRCCDMYTIKNLHAKIIEKDMIKVLOORS 428
 361 RKPSKTEQLSCMRPAKSLMSISNAGSLLSHSTLTGSPIMEEKDDKSKWKSGLGILLG 420
 429 RKDAGKTDS-SSLRPARSVPSIA-AATGTHSRQTSLSLSSQLAEKKEKTKWKSGLGILLG 486
 421 GDYRAEYVPSTPSPVPPSTPL-----SAHSKTSGRDCSTOTERGTE 462
 487 KEHEH--ASAPLLPPPTSALSSTIASTAASSAHAKTSGKSDSTQTDKSAE 536

SULT 7
 C38518

ADC38518 standard; protein; 801 AA.

ADC38518;

18-DEC-2003 (first entry)

Human AMLP1a myosin-tail motif amino acid sequence SEQ ID NO:867.

human; angiometin-like protein 1; AMLP1; cytostatic; gene therapy.

Homo sapiens.

WO2003037931-A2.

08-MAY-2003.

01-NOV-2002; 2002WO-US035129.

01-NOV-2001; 2001US-0334773P.

(AMSH) AMERSHAM BIOSCIENCES SV CORP.

Shannon M, Phan T;

WPI; 2003-430501/40.

New isolated nucleic acid molecule encoding a human angiometin-like protein, useful for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP1.

Example 2; SEQ ID NO 867; 172pp; English.

The present invention describes the human angiometin-like protein 1 (AMLP1). human AMLP1 has cytostatic activity, and can be used in gene therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and compositions of the present invention can be used for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP1. The present sequence represents a myosin-tail motif amino acid sequence, which is used in an example from the present invention.

Sequence 801 AA;

Query Match 43.1%; Score 1441; DB 7; Length 801;

Best Local Similarity 61.8%; Pred. No. 6.9e-82;

Matches 291; Conservative 83; Mismatches 83; Indels 14; Gaps 6;

QY 2 PRAOPSSASYQVPADPAPVSRQAQVVEILSDENRNLROLEGCEYKVARLOKVEIQ 61
 DB 331 FQPPAPASPSQQLGDPAPALVERAQVVEILSDENRNLROLEGCEYKVARLOKVEIQ 330
 QY 62 RVSEAYENLVKSSKREALEKAMENKLEGBIRRMHDFNDRLETLERLETANKQLAEKYEYS 121
 DB 391 RISEAYESLVKSTTKRESLDKAMENKLEGBIRRMHDFNDRLETLERLETANKQLAEKYEYS 450
 QY 122 EDTRKTSIQLEFAKNKESOREKELEAEALATARNSTNEDORRHIEIRDOALSNAQAKVVKLE 181
 DB 451 GD-KAAEGHYASQNKFEKLEKELEMELEAAVRTASEDHRRHIELDOALSNAQAKVVKLE 509
 QY 182 BELKKQYVVDKVMQOALVOLQAAACEKEQLEHRLRTRLEHRLSLRTOQCGNCQPT 241
 DB 510 BELREKQAYVEKVKLQALQALQALQALQALQALQALQALQALQALQALQALQALQALQ 569
 QY 242 NVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVAARDTTV 301
 DB 570 NMPEYNAPALLELVREKEERILALEADMTKWEQKYLEESTIRHFANNAATAAARDTTI 629
 QY 302 ISHSPNTSY-DTALREARIQKEEBEILMANRCLDMGRITKLHAQIIKEDAMIKVLOORS 360
 DB 630 INHNRNGSYGSSLEAHIMWEEBEEVVOANRRCCDMYTIKNLHAKIIEKDMIKVLOORS 689
 QY 361 RKPSKTEQLSCMRPAKSLMSISNAGSLLSHSTLTGSPIMEEKDDKSKWKSGLGILLG 420
 DB 690 RKDAGKTDS-SSLRPARSVPSIA-AATGTHSRQTSLSLSSQLAEKKEKTKWKSGLGILLG 747
 QY 421 GDYRAEYVPSTPSPVPPSTPL-----LSAHSKTSGRDCSTOTERGTE 462
 DB 748 KEHEH-EHASAPLLPPPTSALSSTIASTAASSAHAKTSGKSDSTQTDKSAE 797

RESULT 8
 ADC37654

ID ADC37654 standard; protein; 869 AA.

AC ADC37654;

DT 18-DEC-2003 (first entry)

Human angiometin-like protein 1a SEQ ID NO:3.

human; angiometin-like protein 1; AMLP1; cytostatic; gene therapy; AMLP1a.

OS Homo sapiens.

PN WO2003037931-A2.

PD 08-MAY-2003.

PF 01-NOV-2002; 2002WO-US035129.

PR 01-NOV-2001; 2001US-0334773P.

XX (AMSH) AMERSHAM BIOSCIENCES SV CORP.

PI Shannon M, Phan T;

XX WPI; 2003-430501/40.

DR N-PSDB; ADC38522, ADC38523.

XX New isolated nucleic acid molecule encoding a human angiometin-like protein, useful for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP1.

PS Claim 14; SEQ ID NO 3; 172pp; English.

XX The present invention describes the human angiometin-like protein 1 (AMLP1). human AMLP1 has cytostatic activity, and can be used in gene therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and

compositions of the present invention can be used for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP. The present sequence represents human AMLP from the present invention.

Sequence 869 AA;

Query Match 43.1%; Score 1441; DB 7; Length 869;
Best Local Similarity 61.8%; Pred. No. 7.7e-82;
Matches 291; Conservative 83; Mismatches 83; Indels 14; Gaps 6;

2 PRAPSSASYQVPADPAIVSRQQVVEILSDNRNLROELGCTYKVARLOKVTETIQ 61

331 PQPPAPASQQLGPDFAIVRAQQVVEILTEENVLHQELQCYVDNADLKFKEKLQ 390

62 RVSYAYENVKSSKRALEKAMNKLEGEIRRMHDFNRDLRLELTANKOLAEKEVGS 121

391 RISAYSLVSTTKRASLDKAMNKLEGEIRRMHDFNRDLRLELTANKOLAEKEVGS 450

122 EDTRTISQLFARKESQREKELEAELATARNSTEDQRRHIEIRDQALSNAQAVKLE 191

451 GD-KAAEGHYASQNKFLKEKELEMEELAAVRTASEDHRRHIBILDQALSNAQAVKLE 509

182 EELKKQVYDKVKKVQALVOLQACEKEQLEHRLTSLERLESLRLOPQGNCOPT 241

510 EELREKQVYDKVKKVQALVOLQACEKEQLEHRLTSLERLESLRLOPQGNCOPT 569

242 NVSYNAALMELREKEERILAEADMTKWEQKYLEENVMRHFALDAATAAQQDTTV 301

570 NMPYNAPALLELVREKEERILAEADMTKWEQKYLEENVMRHFALDAATAAQQDTTV 629

302 ISHSPNTSY-DTAEARIQKEEESILMANKRCLEDMGRITKLHAQYIEKDMIKVLQORS 360

630 INHNRNGSYGESSLEAHIMQEEVEVQANRRCCQDMETIKLHAKYIEKDMIKVLQORS 689

361 RKEPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDKSWKSGILIG 420

690 RKDAGKTDSS-LRPARSVPSIA-AATGTHSRQTSLSLQACEKEQLEHRLTSLERLESLRLOPQGNCOPT 747

421 GDYRAEVVPTSPVPSPTL-----LSAHSKTGSDCSTQTERGT 462

748 KEHH-EHASAPLLPPPTSLSSIASTTAASAHAKTSGKSDSTQTDKSAE 797

SULT 9

R56262 standard; protein; 882 AA.

ABR56262;

20-NOV-2003 (first entry)

Mouse Junction-Enriched and Associated Protein, JEAP.

Murine; JEAP; Junction-Enriched and Associated Protein; tight junction; exocrine gland; pancreas; submandibular gland; lacrimal gland; parotid gland; sublingual gland.

Mus musculus.

EPI312615-A2.

21-MAY-2003.

15-NOV-2002; 2002EP-00079841.

16-NOV-2001; 2001JP-00352241.

(EISA) EISAI CO LTD.

Nishimura M, Asano M, Ono Y, Morimoto K, Takeuchi M, Inoue Y, Imai T, Takai Y;

DR WPI: 2003-571281/54.

XX N-PSDB; ACC70742.

PT New junction-enriched and associated protein 2 which is expressed specifically at tight junctions in exocrine glands, useful for identifying substances which affect the expression, activity or distribution of the protein.

XX Claim 13; Page 15-17; 3lpp; English.

XX The present sequence is the protein sequence for murine Junction-Enriched and Associated Protein (JEAP). JEAP is expressed specifically at tight junctions in exocrine glands such as the pancreas, submandibular gland, lacrimal gland, parotid gland and sublingual gland. JEAP is useful for identifying substances which affect its expression, activity or distribution

SQ Sequence 882 AA;

Query Match 41.8%; Score 1400; DB 7; Length 882;
Best Local Similarity 58.2%; Pred. No. 2.9e-79;
Matches 299; Conservative 80; Mismatches 103; Indels 32; Gaps 11;

QY 1 MPRAQP-SSASYQVPADP-----FAIVSRQQVVEILSDNRNLROELGCTYKVA 51

333 LFLPLPISLAASQPLPASPNQQLGPDFAIVRAQQVVEILTEENVLHQELQCYVDNAD 392

QY 52 RLQKVEYTIQVSEAYENVLKVSSKRALEKAMNKLEGEIRRMHDFNRDLRLELTANK 111

393 KLHFKEXELQISBAYESLVKSTTKRESLDKANKTKLEGEIRLHDFNRDLRLELTANK 452

QY 112 QLAKVEYGESEDTKRTISQLFARKESQREKELEAELATARNSTEDQRRHIEIRDQALS 171

453 QLSREYDGHED-KAESHVYVQNKFLKEKELEMEELAAVRTASEDHRRHIBILDQALS 511

QY 172 NAAQKVVKLEELKKVQVYDKVKKVQALVOLQACEKEQLEHRLTSLERLESLRLOPQGNCOPT 231

512 NAAQKVVKLEELKKVQVYDKVKKVQALVOLQACEKEQLEHRLTSLERLESLRLOPQGNCOPT 571

QY 232 QQRQGNCOPTNVSEYNAALMELREKEERILAEADMTKWEQKYLEENVMRHFALDAATAA 291

572 QQKHGTGPPVPSLPECNAPALMELVREKEERILAEADMTKWEQKYLEENVMRHFALDAATAA 631

QY 292 TVAAQRDTTVISHSPNTSY-DTAEARIQKEEESILMANKRCLEDMGRITKLHAQYIEKDMIKVLQORS 350

632 AATAERTTISNHRNGSYGESSLEAHIMQEEVEVQANRRCCQDMETIKLHAKYIEKDMIKVLQORS 691

QY 351 AMIKVLOORSRKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDK- 409

692 AMIKVLOORSRKDAGKTDSS-LRPARSVPSIA-AATGTHSRQTSLSLQACEKEQLEHRLTSLERLESLRLOPQGNCOPT 749

QY 410 SWKSGILIG-----GDYRAEVVPTSPVPSPTL-----LSAHSKTGSDCSTQTERGT 461

750 TWKSGIFLGKEHQOQASAPLPTTPASALSIPASTTSASSTHAKTSGKSDSTQTDKST 809

QY 462 E-----SNKTAAPVAPISVPAPVAAAATAA 485

810 ELFWPSMASLPSRCRLSTAPSNPILKHPAKGA 843

RESULT 10

ABR56263

ID ABR56263 standard; protein; 772 AA.

XX ABR56263;

AC ABR56263;

XX 20-NOV-2003 (first entry)

DE Mouse Junction-Enriched and Associated Protein-2, JEAP-2.

XX Murine; JEAP-2; Junction-Enriched and Associated Protein-2;

KW exocrine gland; pancreas; submandibular gland; lacrimal gland;

KW tight junction.


```

1  Query Match
2  Best Local Similarity 31.1%; Score 1040.5; DB 3; Length 849;
3  Matches 237; Conservative 80; Mismatches 127; Indels 65; Gaps 10;
4
5  13 PVPADPFIVSRAQOMVEI-----LSDENENLRQELGCGYKVARLQKVEITQRV 63
6  PVSQAQSSATSGSAHLAQMEAVLRNARLQDNELQRELESSAKAGRIEKESEIQLRL 424
7
8  64 SEAYENLVKSSKRALKAMRNKLEGIIRNHDNFDRLRLTANKOLAEKYEBSGD 123
9  SEAHESLFRASKEALEKTRNKMDSMRQLQFNDRLESLRLESANRLASKTKTQEAQAG 484
10
11 124 TRKTSISQIFAKNKESORKEKLEAEIATARNEDDORRHLEIRDOALSACAKVKVLEE 183
12
13 485 SQDWAKLLASQYEQOQEKLEEMALLRGAIEDQRRRAELLEQALGNAGRAARABEE 544
14
15 184 LKKQVYVDKVKQQAALVQLQAACEKREQLERLRLTLELESLSRIOQRQ-----GNC 238
16
17 545 LKKQAYYKVERLQQAALGQLQAACEKREQLERLRLTLELESLSRIOQRQAGAPGSS 604
18
19 239 QPTNVSEYNAALMELAEKEERILALEADMKNQKYLEENWHEALDAATVAQRD 298
20
21 605 GSGGSPFLSRLSEQLREKEEQILALEADMKNQKYLEENWHEALDAATVAQRD 664
22
23 299 TTVISHSPNTSYDTALEARIQKEEIEILMANKRCLDVEGRITKLHAQIIIEKDAWKYLQ 358
24
25 665 TTLIRHSPQSPSSSF-----NEGLLTGCHREHESRLKVLHAQILEKDAWKYLQ 717
26
27 359 BSRKPSKTEOLSCMRPAKLSMSINAGSLLSHSSLTITGSPINKEKDDKSKGSLGIL 418
28
29 718 RSRDPGKAIQGS-LRPAKSPVPSFAAAA-----GTQGWGQ-----L 754
30
31 419 LGGYRAEYVPS-----TPSPVPPTPLLSAHSKTSRDCSTQTERGTESNKTAAV 469
32
33 755 SSSBEQTADAPARLTADRAFTPEPVVTAPE-AHAHAKHSRDSGTQDGPDPSTSTCL- 811
34
35 470 APISVPAP-----VNAATAAATTAAT 493
36
37 812 ----PPEPDSLGCSSQRAASLDSVATS 836
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
```

[illegible]

```

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 11448; 2537bp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AA031366 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AA097446 to AA095893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 706 AA;
XX
XX Query Match 29, 58; Score 987.5; DB 4; Length 706;
XX Best Local Similarity 52.38; Pred. No. 1.7e-53;
XX Matches 214; Conservative 65; Mismatches 101; Indels 29; Gaps 5;
XX
QY 13 FVPADPPAIVSRAQOMVEI-----LSDENRNLRQELGECYKVAQLQKVETEIQRV 63
DB ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 296 FVSAQASATSGSAHLAQMBAVLRENARLQDRNRLQRELESSAKAGRIKLESEIQR 355
QY 64 SSAYENLVKSSKREALEKAWRNKLEGEIRRMHDFNRDLRERLETANKOLAEKVEGSD 123
DB ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 356 SEAHESLUTRASKREALEKTRNNKNDSEHARLQDFNRDLRERLESANRELAKTQEAQAG 415
QY 124 TRKTIQLFAKNKESQREKILBAELATARNSTVEDQRRIEIRDQALSNAQKVKLREE 183
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 416 SDQMVAKLIAQSYEQEQEKILREMFALLGATEDQRRRAELIEQALGNAQGRAABEE 475
QY 184 LKKKQVYVYKVKQQAALVOLQACREKQELERLRLTRLEERLESRLRIQDQOQ-----GNC 238
DB ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 476 LKKQAVYKVKERLQQAALGQQAACREKQELERLRLTRLEQELKALRAQORQAGAPGSS 535
QY 239 OPTNVSYNAALMELLRKEERILAEADMTKWQKYLEENVRHFALDAATAATVAAQRD 298
DB ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 536 GSGSGPELSALRLSEQLRKEEQILALEADMTWEQKYLEEHAMRFAMDAATAATAAQD 595
QY 299 TTVTISHSPNTSYDTALEARLQKREERITLMANKCLDMGRIKTLHAQITKXDMATKVLQ 358
DB ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 596 TTLIRHSQPPSPSSP-----NEGTLTGHRRHQWESRLKVLVHAQILLEKDAVILVQ 648
QY 359 RSRKPSKTEQLSCMRPAKSLMISNAGS-----GLLSHSTLTGSP 400
DB ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 649 RSRDPGKAIQGS-LRPAKSPVPVFAAAAAAGTQWGLSSSSSRQTADAP 696
XX
RESULT 14
ABGI7959
ID ABGI7959 standard; protein; 759 AA.
XX
XX ABGI7959;
XX
XX DT 18-FEB-2002 (first entry)

```

RESULT 14
ABG17959
ID ABG1
XX
AC ABG1
XX
DT 18-P

Novel human diagnostic protein #17950.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Dzmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS82146.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 48318; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

Sequence 759 AA;

```
Query Match          27.0%; Score 903; DB 4; Length 759;
Best Local Similarity 99.5%; Pred. No. 3.8e-48;
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1  MPRAQPSASVQPVDPADPPAIVSRAQMVILSDENNLROELGCGYKVARLQKVEYEG 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
525 MPRAQPSASVQPVDPADPPAIVSRAQMVILSDENNLROELGCGYKVARLQKVEYEG 584

61  QRVSEAYENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595 QRVSEAYENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 644

121 SEDTRKITSQIFAKNKESQREKEKLEAELETARSTNEDQRRHIEIRDOALSNAQAKVVKL 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 SEDTRKITSQIFAKNKESQREKEKLEAELETARSTNEDQRRHIEIRDOALSNAQAKVVKL 704

181 EEEL 184
|||||:
```

Db 705 EEV 708

RESULT 15

ADC37657

ID ADC37657 standard; protein; 562 AA.

XX ADC37657;

XX 18-DEC-2003 (first entry)

XX Human angiotensin-like protein 1b SEQ ID NO:6.

XX human; angiotensin-like protein 1; AMLP1; cytostatic; gene therapy;

KW AMLP1b.

XX Homo sapiens.

XX WO2003037931-A2.

XX 08-MAY-2003.

XX 01-NOV-2002; 2002WO-US035129.

XX 01-NOV-2001; 2001US-0334773P.

XX (AMSH) AMERSHAM BIOSCIENCES SV CORP.

XX Shannon M, Phan T;

XX WPI; 2003-430501/40.

XX N-PSDB; ADC37655, ADC37656.

XX New isolated nucleic acid molecule encoding a human angiotensin-like protein, useful for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP1.

XX Claim 14; SEQ ID NO 6; 172pp; English.

XX The present invention describes the human angiotensin-like protein 1 (AMLPI). human AMLPI has cytostatic activity, and can be used in gene therapy. The AMLPI protein, nucleic acid molecules, antibodies, and compositions of the present invention can be used for treating or preventing a disorder associated with decreased or increased expression or activity of AMLPI. The present sequence represents human AMLPIb from the present invention.

XX Sequence 562 AA;

Query Match 22.9%; Score 765.5; DB 7; Length 562;

Best Local Similarity 65.5%; Pred. No. 1.1e-39;

Matches 152; Conservative 39; Mismatches 40; Indels 1; Gaps 1;

QY 2 PRAPSSASVQPVDPADPPAIVSRAQMVILSDENNLROELGCGYKVARLQKVEYEG 61

||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

332 PQQPAAASPQQLGDFADFAIVRAQMVILTBENRVLHOELQGYDNDKLNKFEKELQ 391

||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

62 RVSEAYENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEGS 121

||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

392 RISEAYESLVKSTTKRESLDKAMRNKLEGEIRRLHDNRDLRERLETANKQLSSREYEGH 451

||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

122 EDTKRTISQIFAKNKESQREKEKLEAELETARSTNEDQRRHIEIRDOALSNAQAKVVKLE 181

||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

452 ED-KAAEGHYASQNKFFLKEKELEMELEAAVRTASDEDRHIEIRDOALSNAQAKVVKLE 510

||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

182 EELAKKQVYVDKVKEMQOALVQLQAACEKREQLERLRLRLERLESLRQQ 233

||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

511 EELREKQAYVEKLEKQALQALQLOSACEKEQERRRLRLRLERLESLRQQ 562

||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Search completed: June 16, 2004, 19:12:37

Job time : 88.1687 secs

Sequence 3, Appli
Sequence 31, Appl
Sequence 54, Appl
Sequence 2, Appli
Sequence 2, Appli
Sequence 85, Appl
Sequence 12, Appl
Sequence 1, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 47, Appl
Patent No. 5273901
Patent No. 5482709
Sequence 2, Appli
Sequence 6, Appli
Sequence 6, Appli

28 209.5 6.3 1315 3 US-08-899-595-3
29 209.5 6.3 2662 4 US-09-595-684B-31
30 209 6.2 219 2 US-08-557-309B-54
31 208.5 6.2 1248 2 US-09-080-897-2
32 208.5 6.2 1248 3 US-09-323-735-2
33 207 6.2 803 4 US-09-154-750A-85
34 207 6.2 803 4 US-09-665-479A-12
35 206.5 6.2 1939 4 US-09-310-187A-1
36 205.5 6.1 1886 4 US-08-938-103-3
37 204 6.1 2101 1 US-08-195-487-4
38 204 6.1 2101 5 PCT-US93-06160-4
39 203.5 6.1 550 4 US-09-616-289-47
40 203 6.1 180 6 5273901-7
41 203 6.1 180 6 5482709-6
42 203 6.1 576 2 US-08-533-306A-2
43 203 6.1 576 2 US-08-742-923A-2
44 203 6.1 816 2 US-08-533-306A-6
45 203 6.1 816 2 US-08-742-923A-6

ALIGNMENTS

RESULT 1
US-08-714-741-41
; Sequence 41 Application US/08714741
; Patent No. 6506613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Hart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-714-741-41

Query Match 8.2% Score 275; DB 4; Length 1231;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: June 16, 2004, 19:08:12; Search time 28.0562 Seconds
(without alignments)
1242.060 Million cell updates/sec

le: US-09-332-063-2

fect score: 3347
quence: 1 MPRAQPSASVQVFADPFA.....KTPILQGBDAEMVEYLI 675

ring table: BLOSUM62
Gapop 10.0, Gapext 0.5

arched: 389414 seqs, 51625971 residues

al number of hits satisfying chosen parameters: 389414

imum DB seq length: 0
imum DB seq length: 200000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : Issued Patents 2A:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/POCUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	275	8.2	1231	4	US-08-714-741-41
2	263	7.9	8991	4	US-08-714-741-32
3	252	7.5	2972	4	US-09-579-181-2
4	252	7.5	3118	4	US-09-579-181-1
5	250.5	7.5	399	4	US-09-252-991A-22853
6	237.5	7.1	803	3	US-09-103-429A-4
7	235	7.0	786	3	US-09-103-429A-3
8	234.5	7.0	1326	4	US-09-688-188B-15
9	234.5	7.0	1326	4	US-09-291-417D-15
10	231	6.9	316	4	US-09-252-991A-32957
11	226.5	6.8	2482	1	US-08-328-254-6
12	225.5	6.7	756	4	US-09-963-137-184
13	225	6.7	3248	1	US-08-353-700-1
14	225	6.7	3248	5	PCT-US95-16216-1
15	223.5	6.7	1079	4	US-09-489-039A-7502
16	216	6.5	442	3	US-08-834-306-52
17	216	6.5	442	3	US-08-993-674A-52
18	216	6.5	442	4	US-09-256-976-52
19	213	6.4	262	1	US-08-403-379A-1
20	213	6.4	262	2	US-08-929-414-1
21	213	6.4	263	2	US-08-557-309B-51
22	213	6.4	263	3	US-08-834-306-51
23	213	6.4	263	3	US-08-993-674A-51
24	213	6.4	263	4	US-09-256-976-51
25	213	6.4	1972	4	US-08-875-435B-3
26	212.5	6.3	1274	4	US-09-095-443-2
27	210.5	6.3	1075	4	US-09-252-991A-18387

Best Local Similarity 22.3%; Pred. No. 4.5e-10;
Matches 144; Conservative 94; Mismatches 237; Indels 172; Gaps 24;

22 VSRAQQVVEILSDENRNLRLCEGCGYKVARLQKVEIQRVSEAYENLVKSSSKREALE 81
236 VKAAKDAADKMWIDEAKKREBAKTRFNTVRAMVKEAELELVKEE-----ANESRNEEKI 290
82 KMWKLE---GEIRSMDFNDELRERLETANKOLAEKEYEGSEDTKTIQSOLPAKNKES 138
291 KQKAEVSKVBAEATRLKIKTRKKAEPPEQAETK-KGSEAKQKAPELTKLEEA 349
139 QREKEKLEELATARNSTNEDQRRHIEIRDOALSNAQAVKVLKEELKKQVYVD-----K 193
350 KRKAESSEKAAEAQKQVDAEYVALEAK---IAELEVEQSLKELXE-----IDERAKAK 402
194 VEMQOALVQQAACKEB-----OLEHRLRLERLESLRIQ-----RQGNCPQ 240
403 LEEAEKATFAEKQVDAEYVAQKIAELEQVH-RLSEQLKAGIDESOSDYKEGLRAP 451
241 TNVSEYNAALMELLSEKEERILALEADTKWE-----QKYLENVMRHFDLAAATVA 294
462 LOSKLDTKKALSKLEELSDKIDELDAEIAKLEVLQLSSESDYAKEGP----- 508
295 AQSDTTVISHSPNTSVDTALBARIQKEEELMANKRCLDMEGRIKTLHAQIIEKQAMIK 354
509 -----RAPLOSKLDKAKKSKLEB-----LSDKIDELDAEIAKLEDLQK 548
355 VLQORSRKPSKTEQLSCMRPAKSLMSISNAGSLSHSSTLAGSPIMEKRDDKSKWGS 414
549 DAEGNNVAYFKEGE-----EKTAAEKAELEKAEADLK----- 593
415 LGILLGGVRAEYVSTSPVPTLLSAHSKGTGRDCSTQTERGTESNKTAAPVAPISV 474
584 -----XAVDETPAPAPQK-----AABENNVED-----YKEGLEK----- 616
475 PAPVAAATAAATATATITTTTVAAPVAVAAAAAPAAAAAPSPATAATAAASPAA 534
617 -----TIAAKKAELEKTEADLK-----AVNRPEKPAFAPEPAP-----APFKPAKPA 661
535 AGQIPAAASVAGAAVAPSAASAAAQVAPAPVAPAPVAPAPVAPAPVAPAPVAP 594
662 APEKPA-----APEKPAPEKPAPEPAP-BAPAEQKPAPEPAPAPAPAPAPPEK- 710
595 PPSAPAVAPTAPT-----PTPVAQ-AEVPASPATG 625
711 ----PABQPKPKTKDQQAEBDYARRPEKPAPEKPAPEKPAPEKPAPEKPTPT 753

SULT 2
--08-714-741-32
Sequence 32, Application US/08714741
Patent No. 6500613

GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yocher, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESS: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer Bsg., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 8991 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-32

Query Match 7.9%; Score 263; DB 4; Length 8991;
Best Local Similarity 19.3%; Pred. No. 3.4e-08;
Matches 164; Conservative 83; Mismatches 298; Indels 306; Gaps 21;

QY 2 PRAPFSASQYQVPAOPFAIVSRAQOMVEILSDENRN-----LROELGCGYKVAR 52
S303 PAPKEKPAKPAPEP-----KPELKIIDSSESDYKEGFRAFLQSELD-----KQAK 5352
QY 53 LQKVEITEIQVSEAYENLVKSSSKREALEKAMRNKLEGEIRRMHDFNDRLEHRTANKQ 112
S353 LSKLELSKIDELDAEIAKLEDLQKAAE-----ENNVEDYFKEGLEKTLAAKKA 5403
QY 113 LAKEVEGSDTKTISQISQFPAKNKSOREKLEAEATARNSTNEDQRRHIEIRDOALS 172
S404 ELEKT-----EADLKAVNBEKPAPEPSQPEKPAEAPAPAEQPTPTQPEKPAEQPAPA 5460
QY 173 AQAKVYLEELAKKQVYVDKVEKQOALVLOAACEKKEQLEHRLRLERLESLRIQ 232
S461 PQ-----PEKPAETPAPKPEKPAE---QPKAKPADQQAEBDYARRSSEYVNRLLTQ 5510
QY 233 QROGNCOPT-----NVSEYNAALMELLSEKEERILALEADTKWEQKYLEENVMRH 284
S511 QPPKAEKPAPEKTKGSGALDQEAAPPHQVADLEKQITGPEIFLG----- 5556
QY 285 PALDAATVAAQORDTTVISHSPNTSYDTALBARIQKEEELMANKRCLDMEGRIKTLHA 344
S557 -GADPEADTAARENSLAQKAEALQKPTGLEKLLDS-----LDPGKTKQ----- 5599
QY 345 QIIEKQAMIKVLOORSKEPSKTEQLSCMRPAKSLMSISN-----AGSGLSHSSTLIGS 399
S600 DELDKAGEBAELDKKADLPKNVADLE-----KEISNLEILLGGADSEDDTAALPNK 5651
QY 400 PIMEEKRDDKSKK---CSGLILGSGDYARYVPS----- 430
S652 LAXKQAELEKTKQELDAENELGPDGDEETTPAPAPQEPQAPAPKPEQAPAPKPEQAPA 5711
QY 431 -----TPSEVP-----PSTPL----- 441
S712 PAPKEQAPAPAPKPEQAPAPAPKPEQAPKPEABEPTQPEKPAETKTRVRLKVAEFGVQ 5771
QY 442 -----LSHSHKTSRDCSTQTERG-----TESNKTAAVAPISVAPVAAA 481
S772 LRDAGSNNVAYFKEGLEETTAEXEAGLGKAEADLKAVDEPETPAPAPAPAPAPAPAP 5831
QY 482 ATAAATATAATITTTTVAAPVAVAAAAAPAAAAAPSPATAATAAASVAPAAAGQIPAA 541
S832 APKPA-----PAPKPAPEAPAPAPAPKPAPEKPAPEKPAPEKPAPEKPAPEK 5878
QY 542 A----- 542

5879 APKPTPKTKLIDESDEYAKELGRAPLQSELDTKKAKLLKLELSEKIELDAEIXE 5938
543 -----SVASARAVAPSAASAAAV 560
5939 LEVOLDKDAEGNNVVEAYFKEGLEKTTAEKKAELKABDLKKAVDZPETPAPAPAPAP 5998
561 QVAPAPAPAPAPALVEVPAPAPAAQASAPAQTOAPTSAPAVAPTPTPTTAPVAAQAEVPA 620
5999 APTPEAP 6058
621 -SPATGCPHR 630
6059 PAPAPAPKPKX 6069

SULT 3

-09-579-181-2
Sequence 2, Application US/09579181
Patent No. 6365372
GENERAL INFORMATION:
APPLICANT: Chiviva, John
APPLICANT: Yaciuk, Peter
TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
FILE REFERENCE: 16153-4247
CURRENT APPLICATION NUMBER: US/09/579,181
CURRENT FILING DATE: 2000-05-25
PRIOR FILING DATE: 2000-05-25
PRIOR FILING DATE: 1999-05-27
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 2972
TYPE: PRT
ORGANISM: Human
-09-579-181-2

Query Match 7.5%; Score 252; DB 4; Length 2972;
Best Local Similarity 30.5%; Pred. No. 4.6e-08;
Matches 98; Conservative 34; Mismatches 103; Indels 86; Gaps 15;

430 STSPVPVPSTPLLSAHSKTSRDCSTQTERGTSNKTA-AVAPISVP-----475
1153 SSPWPIPNSSPLASPVSVSTVPLSSSLPISVPTTLPAAPASAPLTIPISAPLTVSASGPA 1212
476 -----APVAAA-----TAAAIT---ATAATITTTWAAAPVAVA---507
1213 LLTSVTPLAPVWPAAPGPPSLQPSGASPSASALTGLATAPSLSSQTPGHPHLLAPTS 1272
508 -----AAAAAPAAAAAPSPATAATAAAVSPAAAGQIPAAASVASAAAVAPSAASAAAV 560
1273 SHVPLNSTVAPACSPVLVPASALA-----SPFSPAPNAPAPQ-ASLLAPASSASQALAT 1326
561 QVAP-AAP-----APVPAPALVPV-----APAAAQASAPAQTOAPTSAP-----AV 601
1327 PLAPWAAPQTAILAPSPAPPLAPLPLVAPSGAAPVLASSQTPVFNAPSTFTGTSLSA 1386
602 APTPAPTPT--TPAVAQAEVPASPATGPGPHRLSIPSLTGNP--DKTDGVPVFNHNTLERKT 657
1387 SPVPAPTPLVAFSSQTQMLPA-PVPSELPSPASTQTLLALAPALAPTLLGGSSPSQTLISLT 1445
658 -----PIQLGQEPDAEMV 671
1446 GNPQGPPTQTLSLTPASSLV 1466

RESULT 4

-09-579-181-1
Sequence 1, Application US/09579181
Patent No. 6365372
GENERAL INFORMATION:
APPLICANT: Chiviva, John
APPLICANT: Yaciuk, Peter

FILE REFERENCE: 16153-4247
CURRENT APPLICATION NUMBER: US/09/579,181
CURRENT FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: 60/136,620
PRIOR FILING DATE: 1999-05-27
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 3118
TYPE: PRT
ORGANISM: Human
US-09-579-181-1

Query Match 7.5%; Score 252; DB 4; Length 3118;
Best Local Similarity 30.5%; Pred. No. 4.9e-08;
Matches 98; Conservative 34; Mismatches 103; Indels 86; Gaps 15;

430 STSPVPVPSTPLLSAHSKTSRDCSTQTERGTSNKTA-AVAPISVP-----475
1299 SSPWPIPNSSPLASPVSVSTVPLSSSLPISVPTTLPAAPASAPLTIPISAPLTVSASGPA 1358
476 -----APVAAA-----TAAAIT---ATAATITTTWAAAPVAVA---507
1359 LLTSVTPLAPVWPAAPGPPSLQPSGASPSASALTGLATAPSLSSQTPGHPHLLAPTS 1418
508 -----AAAAAPAAAAAPSPATAATAAAVSPAAAGQIPAAASVASAAAVAPSAASAAAV 560
1419 SHVPLNSTVAPACSPVLVPASALA-----SPFSPAPNAPAPQ-ASLLAPASSASQALAT 1472
561 QVAP-AAP-----APVPAPALVPV-----APAAAQASAPAQTOAPTSAP-----AV 601
1473 PLAPWAAPQTAILAPSPAPPLAPLPLVAPSGAAPVLASSQTPVFNAPSTFTGTSLSA 1532
602 APTPAPTPT--TPAVAQAEVPASPATGPGPHRLSIPSLTGNP--DKTDGVPVFNHNTLERKT 657
1533 SPVPAPTPLVAFSSQTQMLPA-PVPSELPSPASTQTLLALAPALAPTLLGGSSPSQTLISLT 1591
658 -----PIQLGQEPDAEMV 671
1592 GNPQGPPTQTLSLTPASSLV 1612

RESULT 5

US-09-252-991A-22853
Sequence 22853, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 22853
LENGTH: 399
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22853

Query Match 7.5%; Score 250.5; DB 4; Length 399;
Best Local Similarity 25.4%; Pred. No. 4.6e-09;
Matches 107; Conservative 42; Mismatches 170; Indels 103; Gaps 13;

212 EOLHEHLATRLERSLESRIQORQNCQPTNVSYNAALMELLREKEERILALEADMTK 271
63 OOLSHSL-----VEHLGACK-----QALVDSEKULAKLEKQK 97


```

399  ---NQALRREERQA-----EREQEVIRHLEES-----ORQLEI-----431
307 NTSYDTALEARIKEREHEELMANKRCLDMEGRLKTLHAQILEXDAIKVLOQRSRKBPSS 366
432  ---LOQALLQELLEYKRLQEEORQERQORLQOQHAYLKSLQOQOQOQLO 484
367 TEQLSCMRPAKSLMSISNAGSLHSSHTLTGSPIMEEKRDDKSWKSGSLGILGGDYRAE 426
485 KQOQOQLLPG-DRKPLVHYGRGMNADKPAWAZEREVERTRMNKQONSFLAKSKPGSTGPE 543
427 YVSTP--SPVPSSTLLSAHSTGSDSCSTOTERTGESNKTAAVAPISVPAPVAAAATA 484
544  --PPIQASPGPG-PL-----SOTPMQRVPEQEGPHKSLVHVRPLK---PYAAPVPR 593
485 AAITATAATITTTWAAAPVAVAAAAAPAAAAAPSPATAATAATAAVSPAAGQIPAAASV 544
594  --SQSLQDQOTRNLAAPP-----ASHDPDPAIPAPTA---TPSARGAVIRQNSD 637
545 ASAAAVAPSAAAAAAVQVAPAAPV--APALVPVPAPAAAAQASAPAOQTAPTSAPAVA 602
638 PTBEGPGFSPNPAWVRPDNEAPKVPQRTSGTSIATALTNTGAGSRRAQ-----AVR 689
603 PTAP-----TPTPAVAQAVPASPATGPGPHRLSISPLTCNPD 641
690 ARPSNSAWOILYORAEERTGPK-----PGPPAOPPGP-----PNASNPND 731

```

```

RESULT 9
US-09-291-417D-15
; Sequence 15, Application US/09291417D
; Patent No. 6680170
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT APPLICATION NUMBER: US/09/291.417D
; CURRENT FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1326
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-291-417D-15

```

	Query Match	7.0%;	Score 234.5;	DB 4;	Length 1326;
	Best Local Similarity	21.7%;	Fred. No. 2.4e-07;		
	Matches 129;	Conservative 99;	Mismatches 237;	Indels 129;	Gaps 24
QY	68	ENLVKSSKR--EALKAMRNKLEGEIIRMHDFNRDLRRLETANKLQAERKEVSGESDTRK	126		
DB	247	EQLLKFPPIRDOPTIERVRIQLKHID-----RSKKRGEXEEETEVYSGSE---	294		
QY	127	TISQLFARKNSQRBEKLEAEALATASTNEDQRHHIEIRDQALSNAOKVKVLEEELKK	186		
DB	295	-----EDDSHGEEGPSSIMNVPGESTIRREFLRLOEQNKENSEA--LKQQQLQQ	343		
QY	167	KQVVYDKVKMQVALVOLQAACEKREQEHLRTLTRELESRLRIQQQGNCQPTNVSEY	246		
DB	344	QQ-QEDPEAHIKHLLHQQRRIIEQEKERRAVEEQORREREQRKLQEKB---QORRLD-	398		
QY	247	NNAALMELLRKEBERILALEADMTWKEGKYILEENVVRHFALDAATAVAQRDTTVISHSP	306		
DB	399	-----MQALRESESRQA-----EREQYEIRHRLEEE-----QRGLEI-----	431		
QY	307	NTSYDTALSAIRQEEBEEBILMANRCCLDMGRIRITLHAQIIHKDAMIKVLAQRSRKBSK	366		
DB	432	-----LCOOLQSFQALIAFYRKQLSEORSERLQSLQOEHAFLASLQSQSQSQSQQLQ	484		

367 TEQLSCMRPAKSLMSISNAGSLSSHTLTGSPIMEEKEDKSWKSGILGILLGGDYRAE 426
485 KQOQQQLPG-DRKPLTHYGRGNPKADKPAWAVEVERBTRMNNQOQSPAKSGPGTGF 543
427 YVPSTP--SPVPSPTPLSAHSGTGRDCSTQTERGTESKNTAAVAPISVPAPVAAAATA 484
544 --PPIQASGPPG-PL-----SQTTPMQRVPBPQEGPHKSLVAHVRFLK---PYAAPVR 593
485 AAIATATAATITTTMVAAPVAVAAAPAAAPAAAPSPATAATAAAVSPAAAGOIIPAAASV 544
594 ---SQSLQDOQTRNLAAFP-----ASHDDPAIPAPTA---TPSARGAVIRQNSD 637
545 ASAAVAPSAASAAAQVAPAAAPVP--APALVPVPAPAAQAASAPAOQTOAPTSAPAVA 602
638 PTSGPGSPNPPAWREDNEAPKVPQRTSSIIATALTSGAGGSRPAQ-----AVR 689
603 PTPAP-----TPPTAVAAQAEVPASPATGPGPHLSIPSLTCHPD 641
690 APRSRNSAWQIYLORRAERTGPKD-----PGPPAQPQGP-----ENASSNPD 731

SULT 10
-09-252-991A-32957
Sequence 32957, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196 136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32957
LENGTH: 316
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
-09-252-991A-32957

Query Match 6.9%; Score 231; DB 4; Length 316;
Best Local Similarity 29.2%; Pred. No. 6.7e-08;
Matches 107; Conservative 29; Mismatches 129; Indels 102; Gaps 15;
302 ISHSPNTSYDTALEARIQKBEIEILMANKRCLDMEGRIKTLHAQI-----TEKDA 351
6 VSRBNKESRMAGKKSEKSSWI-----GEIEKYSRQIMLAGLGAYSKVSXDG 54
352 MIKVLQORSRKEPKTEQLSCMRPAKSLMSISNAGSLSSHTLTGSPIMEEKRD--- 407
55 --SKLFETLVLDGEKAE-----KEAKSDVDQVGAAKASARSASKV-DEVRDALG 103
408 -----DKSWKSGILGIDGYRAEYVBPSTPSPVPTPLLSAHSKTSRDCSTQTE 458
104 KXSELEAEAFDXLNSAISL-----GVPSRNEVKELHSHKVDI--LTKQIE 146
459 RGT-----ESNKAATAVAISVPAPVAAATAATAATAAATITTT 497
147 KLTGVSVKPAKAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 203
498 MVAAAPVAAAPVAAAPVAAAPVAAAPVAAAPVAAAPVAAAPVAAAPVAAAPV 557
204 KTAAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 260
558 AAQVAPAPAPVAPALVPVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 617
261 A-----KPAKKPA-----AKKPAKPAKPAKPAKPAKPAKPAKPAKPAK 307
618 VPASPAT 624
|||:::

Db 308 APATPSS 314
RESULT 11
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-bwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CU 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 6.8%; Score 226.5; DB 1; Length 2482;
Best Local Similarity 21.4%; Pred. No. 1.8e-06;
Matches 139; Conservative 100; Mismatches 233; Indels 179; Gaps 26;
QY 26 QQVVEILSDENRNLRLQELE-----GCYEKVARLOKVETIORVSEAVENLVKSSKRE 78
1948 QDTLEVLSQSSYKLNLENELELTQKDKMSFVEKNKMTAKETELQ--EMHEMAQKTAELQE 2005
QY 79 ALEKAMNKLEGEIRMHED-----FNRLRERLETANKOLAEKEYEGSED- 123
2006 EL-SGEKNRLAGELQLLEBTKSKKOLKELTLENSELKSLDKMDKQVKEKGVSRSEI 2064
QY 124 -----TRKTIISOLFANKEQREKLEAEALATARSTNEDQRRHIEI---RDQAL 170
2065 AEYQLRLHAEKKGQALLDNTNKQYVEIQTYYREKLTSECECLSSQKLEIDLLKSSKEEL 2124
QY 171 SNAQKVVVLEBEELKKQV-----YVDKV-----EKMQQALVOLQAAACEKBPOLSHRLTRL 222
2125 NNSLKATTTQLEELKTKMDNLKYNQKXENRAQGMKLLKSKNQLEEBKEILOKEL 2184
QY 223 ERELESRIQORQNCQOPTNVSEYNA--AALMELLREKEERILALEADMTKWSQKYLEEN 280
2185 S-QQAAQEKQKTQVMDYKVDLTTEIKELKELEK-----TKBADEYLDKY 2232
QY 281 VRRHFALDAATAVAQRTDVTTVISHSPNTSYDTALEARIQKBEIEILMANKRCLDMEGRIK 340
2233 C-----SLLISHE-----KLEKAKEM-----LETOVA 2254


```

176 KVKLEBELKKQVYVVKVQKQALVQQAACEKEQLEHLRTRLERELESRIQORQ 235
674 -----VRQTSGDDEAK-----AKSRDEQOPRRRTA--RRSDDKROAQOE 711
236 GNCQPTN---VSEYNAALMELLREKEERILALEADMTKEOKYLEENVRHFALDAAAT 292
712 AKACTREEPVQOE-----TEQEEVOTL-----PRKPRQLAQKVR----- 747
293 VAAQRDTTVISHSPNTSYDTALEARIQKEBEEILMANKRCLDMEGRIKTLHAQIIEKDAM 352
748 -----VESAVVEPVAEIIVPEAVV 765
353 IKVLQORSRKEPSKTLQSCMRPAKSLMSISNAGSLLSHSSTLTGSPIMEKRDOKSWK 412
766 ABVIAPES--EPVKAE-----LPA-GVESVADQDEN--GESREANGMP-----RRSRR 808
413 GSLGILLGSD---YBAEYVPSTPSPVPSTPLLSAHSKTGS-----RDCSTQTRBG 460
809 SPRHLRVSGQRRRYRDERYP-TQSPMLUTVACASPEMASGKWIRYPVWRPQDOQPBEV 867
461 TESNKTAAVAPISVPAPVAAAATAAATAATITTTMVAAPVAVAAAAAPAAAAAPSP 520
868 QVQDASVAKTVEAAPVAVVETVTAAPVTVEPATMEPVTAEPVVV---EPVAAAEPPLF 923
521 ATAA--ATAAASPAAGQIPADAASVASAASAAAQVAPAPAPVPAPAL-VP 577
924 VDAAEVVAAPAAVEPAPOEPVTEAPAVEAFOAIAPVTLDAEPVVVEPEAVETTPVVAAPVE 983
578 VPAPAAQAASAPQTOAPTSAAPAVAPTPTPTPAVAQAEVPASPATGPGP 628
984 TIAPVAETVEQAPVTEAAPAEPVKAEPVSKPVVVAHERHATAPMTRAPAP 1034

```

arch completed: June 16, 2004, 19:13:23
time : 31.0562 secs

APPLICANT: LINCOLN, Stephen E.
TITLE OF INVENTION: SECRETORY MOLECULES
FILE REFERENCE: PT-1134 PCT
CURRENT APPLICATION NUMBER: US/10/204,887
CURRENT FILING DATE: 2002-08-21
PRIOR APPLICATION NUMBER: 60/185,215; 60/185,216; 60/205,232; 60/205,323; 60/205,287;
60/205,324; 60/205,286
PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
2000-05-17; 2000-05-17
NUMBER OF SEQ ID NOS: 159
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 608
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 18, 388
OTHER INFORMATION: unknown or other
-10-204-887-87

Query Match 43.2%; Score 1447; DB 14; Length 608;
Best Local Similarity 62.1%; Pred. No. 2.4e-74;
Matches 293; Conservative 79; Mismatches 84; Indels 16; Gaps 6;
2 PRAPSSASQVPADPFAIVSRAQQVWELSDENRNLRQELGGCYEKVARLQKVETEIQ 61
70 POPPPASPOQLGPDFAFAIVERAQQWVILTEENVLHQBQLGQYDNDKHLKEKELQ 129
62 RVSEAYENLVKSSSKREALEKAMRNKLEGBIRRMHDFNDRLERLSTANKQALAEKEYEGS 121
130 RISEAYESLVKSTTKRESLDKAMRNKLEGBIRRMHDFNDRLERLSTANKQALAEKEYEGS 189
122 EDTRKTSIQFPAKNSQREKELEALATARNSTNEDORHIEIRDOALSNAQAKVVKLE 181
190 ED-KAAEGHVASQNFELKEKELEALATARNSTNEDORHIEIRDOALSNAQAKVVKLE 248
182 EELKKQVYVDKVKVQQAALVQQAACEKEQLEHLRLTLRLERLESLRIQORGNCOPT 241
249 EELREKQAVYEVKEKQQAALVQQAACEKEQLEHLRLTLRLERLESLRIQORGNCOPT 308
242 NVSEYNAALMELREKEBERILAEADMTKWEQKYLEENVMRHFDALDAATVAARDTTV 301
309 NMPEYNAPALLELVREKEBERILAEADMTKWEQKYLEENVMRHFDALDAATVAARDTTI 368
302 ISHSPNTSY-DTAEARIQKEBERILMANKRCLDMEGRITKTLHAQIIIEKDAMIKVLQORS 360
359 INHSPNGSYGESSLEAHNWEKEBEVQANRRCCODMEYTKNLHAKIIIEKDAMIKVLQORS 428
361 RXEPEKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKDDKSWKSGILILG 420
429 RKDAGKTDS-SSLRPARSVPSIA-AATGTHSRQTSLSSTSSQLAEKKEBKTKWKSIGLILG 486
421 GDYRAEYVPSPTSPVPSPTPL-----SAHSGTGRDCSTOTERGT 462
487 KEHHEH--ASAPLLPPPTLSALSIASHTTAASSAHAKTSGSKDSSTQIDKSAE 536

APPLICANT: Yoshimi Takai
TITLE OF INVENTION: Exocrine gland tight junction-constituting protein JEAP family
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/298,417
CURRENT FILING DATE: 2003-01-22
PRIOR APPLICATION NUMBER: JP 2001-352241
PRIOR FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 882
TYPE: PRT
ORGANISM: Mus musculus
US-10-298-417-2

Query Match 41.8%; Score 1400; DB 14; Length 882;
Best Local Similarity 58.2%; Pred. No. 1.8e-71;
Matches 298; Conservative 80; Mismatches 103; Indels 32; Gaps 11;
QY 1 MPRAQP-SSASYQVPADP-----FAIVSRAQQVWELSDENRNLRQELGGCYEKVA 51
DB 333 LPLPLPISLAASQPLPASPNQQLGPDFAFAIVERAQQWVILTEENVLHQBQLGQYDND 392
QY 52 RLQKVETEIQRVSEAYENLVKSSSKREALEKAMRNKLEGBIRRMHDFNDRLERLSTANK 111
DB 393 KHKFEKELQSISEAYESLVKSTTKRESLDKAMRNKLEGBIRRMHDFNDRLERLSTANK 452
QY 112 QLAKEYEGSEDTRKTSIQFPAKNSQREKELEALATARNSTNEDORHIEIRDOALS 171
DB 453 QLSREYDGHED-KAAESHVYVSQNFELKEKELEALATARNSTNEDORHIEIRDOALS 511
QY 172 NAGAKVVKLEELKKQVYVDKVKVQQAALVQQAACEKEQLEHLRLTLRLERLESLRI 231
DB 512 NAGARVFKLEELREKQAVYEVKEKQQAALVQQAACEKEQLEHLRLTLRLERLESLRI 571
QY 232 QORGNCOPTNVSEYNAALMELREKEBERILAEADMTKWEQKYLEENVMRHFDALDA 291
DB 572 QOKHGTGPPVSLPECNAPALMELVREKEBERILAEADMTKWEQKYLEENVMRHFDALDA 631
QY 292 TVAQRDTTVISHPNTSY-DTAEARIQKEBERILMANKRCLDMEGRITKTLHAQIIIEKD 350
DB 632 AATAERTTISNHRNGSYGESSLEAHNWEKEBEVQANRRCCODMEYTKNLHAKIIIEKD 691
QY 351 AMIKVLQORSKPEKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKDDK- 409
DB 692 AMIKVLQORSKDGKTDAS-LRPARSVPSIA-AATGTHSRQTSLSSTSSQLAEKKEBK 749
QY 410 SWKSGILILG-----GDYRAEYVPSPTSPVPSPTPLLSA-HSKTGRDCSTOTERGT 461
DB 750 TWKSGIFLLGKEHQQAASAPLLPTTPASALSUPASTTSASSTHAKTSGKDSSTQIDKST 809
QY 462 E-----SNKTAAPVAPISVPVAAAAATAA 485
DB 810 ELFWPSNAPSLPSRGLSTAFSPNSPILKHPAAKGA 843

RESULT 3
US-10-298-417-4
Sequence 4, Application US/10298417
Publication No. US20030124603A1
GENERAL INFORMATION:
APPLICANT: Miyuki Nishimura
APPLICANT: Mayumi Asano
APPLICANT: Yuichi Ono
APPLICANT: Koji Morimoto
APPLICANT: Masakazu Takeuchi
APPLICANT: Yoko Inoue
APPLICANT: Toshio Imai
TITLE OF INVENTION: Exocrine gland tight junction-constituting protein JEAP family
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/298,417
CURRENT FILING DATE: 2003-01-22

2709 ACTACAAAAA..... 2768
 553 SAAATAAQAQVAPAPVAPALVPVAPAPAAQASAPACTOAPTSPAPVAPTPTPTPA 612
 2769 AA 2828
 613 VAQAEVPASPA 623
 2829 AAAAAAAAAAAA 2839

SULT 10
 -10-063-685-37
 Sequence 37, Application US/10063685
 Publication No. US20030180909A1
 GENERAL INFORMATION:
 APPLICANT: Eaton Dan L.
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Grimaldi, Christopher J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Watanabe, Colin K.
 APPLICANT: Wood, William I.
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3230R1C1
 CURRENT APPLICATION NUMBER: US/10/063.685
 CURRENT FILING DATE: 2002-05-08
 Prior Application removed - See Palm or File Wrapper
 NUMBER OF SEQ ID NOS: 170
 SEQ ID NO 37
 LENGTH: 2846
 TYPE: DNA
 ORGANISM: Homo Sapien
 -10-063-685-37

Query Match 7.9%; Score 265.5; DB 14; Length 2846;
 Best Local Similarity 42.4%; Pred. No. 2.7e-06;
 Matches 81; Conservative 11; Mismatches 84; Indels 15; Gaps 2;
 448 TGRDCSTQT-----ERTESNKTAATAVAPISVPAPVAAATAA-----AATATAA 492
 2649 TGGTCCTTTCTCTCCCATCTCTGTACACATTTTAAATAATAAGGTTGGCTTCTGA 2708
 493 TTTTWWAAQVAVAAAPAAAAAAPSATAATAAVSPAAGQIPAAASVASAAPV 552
 2709 ACTACAAAAA..... 2768
 553 SAAATAAQAQVAPAPVAPALVPVAPAPAAQASAPACTOAPTSPAPVAPTPTPTPA 612
 2769 AA 2828
 613 VAQAEVPASPA 623
 2829 AAAAAAAAAAAA 2839

RESULT 11
 -10-184-644-149
 Sequence 149, Application US/10184644
 Publication No. US2003004930A1
 GENERAL INFORMATION:
 APPLICANT: Baker, Kevin P.
 APPLICANT: Chen, Jian
 APPLICANT: Desnoyers, Luc
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Pan, James
 APPLICANT: Smith, Victoria
 APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3430R1C27
 CURRENT APPLICATION NUMBER: US/10/184.644
 CURRENT FILING DATE: 2002-06-28
 Prior Application removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 612
 SEQ ID NO 149
 LENGTH: 2773
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-10-184-644-149

Query Match 7.9%; Score 264.5; DB 14; Length 2773;
 Best Local Similarity 43.4%; Pred. No. 3e-06;
 Matches 79; Conservative 11; Mismatches 85; Indels 7; Gaps 2;
 QY 449 GSRDCSTQTERGTESNKTAATAVAPISVPAPVAAATAAATAATAATTTTWWAA-----AP 503
 DB 2594 GTTTCATTTTGTCA--TGCAATGTAGGATTCGCAATTAATGTTTAGAGGATGAA 2651
 QY 504 VAVAAAAAPAAAAAPSATAATAAVSPAAGQIPAAASVASAAPVASAAPVAAVQVVA 563
 DB 2652 AATAAA 2711
 QY 564 PAAPAPVAPALVPVAPAPAAQASAPACTOAPTSPAPVAPTPTPTPAVAPVAPASPA 623
 DB 2712 AA 2771
 QY 624 TG 625
 DB 2772 AG 2773

RESULT 12
 US-10-184-634-149
 Sequence 149, Application US/10184634
 Publication No. US2003006868A1
 GENERAL INFORMATION:
 APPLICANT: Baker, Kevin P.
 APPLICANT: Chen, Jian
 APPLICANT: Desnoyers, Luc
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Pan, James
 APPLICANT: Smith, Victoria
 APPLICANT: Watanabe, Colin K.
 APPLICANT: Wood, William I.
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3430R1C27
 CURRENT APPLICATION NUMBER: US/10/184.634
 CURRENT FILING DATE: 2002-06-28
 Prior Application removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 612
 SEQ ID NO 149
 LENGTH: 2773
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-10-184-634-149

Query Match 7.9%; Score 264.5; DB 14; Length 2773;
 Best Local Similarity 43.4%; Pred. No. 3e-06;
 Matches 79; Conservative 11; Mismatches 85; Indels 7; Gaps 2;
 QY 449 GSRDCSTQTERGTESNKTAATAVAPISVPAPVAAATAAATAATAATTTTWWAA-----AP 503
 DB 2594 GTTTCATTTTGTCA--TGCAATGTAGGATTCGCAATTAATGTTTAGAGGATGAA 2651

504 VAVAAAAAPAAAAAPSPATATAAATAAVSPAAAGQIPAAASVAAAVAPSPAAAAAQA 563
2652 AAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 2711
564 PAAPAPVPALVVPVAPAAQAASAPAOQAPTSAPAVAPTPTPTPAVAQAEVPSA 623
2712 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2771
624 TG 625
2772 AG 2773

SULT 13
-10-063-685-33
Sequence 33, Application US/10063685
Publication No. US20030180909A1
GENERAL INFORMATION:
APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063,685
CURRENT FILING DATE: 2002-05-08
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 33
LENGTH: 2773
TYPE: DNA
ORGANISM: Homo Sapien
-10-063-685-33

Query Match 7.9% Score 264.5; DB 14; Length 2773;
Best Local Similarity 43.48; Pred. No. 3e-06;
Matches 79; Conservative 11; Mismatches 85; Indels 7; Gaps 2;
449 GSRDCSTQTERGTSNKTAVAPIVSPAPVAAAAATAAATATATATTTTAA 503
2594 GTTTCATTTTGTCA--TGACAAATGTAGGAATGCTGAATTAATCTTTAGAGGATGAA 2651
504 VAVAAAAAPAAAAAPSPATATAAATAAVSPAAAGQIPAAASVAAAVAPSPAAAAAQA 563
2652 AAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 2711
564 PAAPAPVPALVVPVAPAAQAASAPAOQAPTSAPAVAPTPTPTPAVAQAEVPSA 623
2712 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2771
624 TG 625
2772 AG 2773

SULT 14
-10-369-493-3279
Sequence 3279, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 3279
LENGTH: 1965
TYPE: PRT
ORGANISM: Neurospora crassa
US-10-369-493-3279

Query Match 7.9% Score 263.5; DB 15; Length 1965;
Best Local Similarity 22.9%; Pred. No. 2.3e-06;
Matches 189; Conservative 110; Mismatches 280; Indels 245; Gaps 38;
26 QQWVEI-----LSDENENLROEGCYKVAELQKVEIQRVSEAYENLVKSSSKREA 79
1038 QETAETKAREDDUTAOQKLLHQOLDGVTQISALAKOSRSAMDESAG-----TOISANDTA 1093
80 LE--KAMRNKL--EGEI-----RRMHDNDRDLRERLETANKQLAEKEYEGSEDTRK--- 126
1094 IEGLELNLYLRREKEILEVQFDLKVQEAR--LQQCLEYSQSOLDARLKLQDERRSQAD 1152
127 -----TISQIFAK-----NKESOREKEKLEAEIATARSTNEDQORRHEIRQOAL 170
1153 SARNSLTHKELDKLNLNLIRESNVLRLNENLRIQALAMKRNKTIEDLENRIQPLEARI 1212
171 SN-----AAQKVKLEEB---LKKK-----QVYVDKVERMQOALVOLQAAACEKR 211
1213 SELELDKSPKEAEVKQLOEARDGLQKRISSILSKYQADPOVEQLKASVESIQAREAF 1272
212 EQLEHRLRTEL---ERELESLR-----IQROGNCQPTNVSEYNA--AALME 253
1273 KQTEAALKEELKEAETNLETERTNWKNMRLAEDPKTRFGNVK--TORNELAAEKQALQT 1331
254 LLREKEERILALRADMTKWEQ-----KYLEENVMRHPALDAAATVAAQDRTTVISHSPNT 308
1332 FLDAANERLSAVEKDLSESARQDLANLOAQNTLOEQHQAQAAAAAASAEQTPVAPVAA 1391
309 S---YDTALEARIQ--KBESEILMANKRCIDMBGRITKTHAQ-----IIEKD-AMIKVLQ 358
1392 APAEQDPLLQOLDALRQELSVKSKAALETE--LESKAEIATAIABRDEARAEVARL 1449
359 RSRKEPKTQLSCMRPAK--SLMSISNAGSGLLSHSSTLTGSP-----401
1450 QSSTHTGDTTEMQDVEAPAPASAPAPAEQGTGLSDBERKALBERIAAAEAKAAEFKAK 1509
402 -MBEKRD-----DKSMKSGILGCGD-----YRAYE-----427
1510 DLEERADAIYKQSEKMKTKALNELAESKEAMEKQTDQERQKLAQAEVDLKLQCELAILKA 1569
428 -----VPSTPSPVP-----PSTPLLSSHSKTSGRD--CSTQTERG---460
1570 EQQSGVSGNGVPATPAKPPPTAQTGAGTTPGLPLANLTDQOTRELISSNQVIMGIK 1629
461 -----TESNKTAAPVAPISVPAP-----VAAAAATAAJTATAATITTTMV-----499
1630 SNVKNTAASKKVRRELEVVKAEYEQKIASKEQAALTEKKSALELNMLORQLKTEK 1689
500 -AAAPVAVAAAAAP-----AAAAASPATAATAAAVSPAAAGQIPAAASVA 545
1690 AKTAVVETAARETPOKPVWVWVWVAKPAAPPPPPAPAT-----LPAPLSRS 1738
546 SAAVAPS-----AAAAAQAQVAPAPAPVAPALVVPVAPAAAAQASA-----588
1739 QRARKSMTKXTNDGSDAAAVVVKP--AGEEIKQIQTPTPTNKAAAPAROGSIVGGQVVG 1797
589 -----PAQTQAP--TSAPAVAPTPTPTTTPAVAAQAEVPSATGP 626
1798 NPFQPGANQSPVTSQPPANPFGAAPQOQQAQPPQF--DATNP 1839

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

n on: June 16, 2004, 19:09:47 ; Search time 27.2311 Seconds
(without alignments)
2384.384 Million cell updates/sec

US-09-332-063-2

Effect score: 3347
Sequence: 1 MPRAQSSASVQVPADPFA.....KTPIQILQEPDAEMVEVLI 675

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

tal number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase : PIR 78:.*
1: Piri:.*
2: Piri:.*
3: Piri:.*
4: Piri:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	300	9.0	1087	1 QPM5H	neurofilament trip
2	292	8.7	507	2 T44768	antifreeze glycope
3	286.5	8.6	2346	2 T13629	Tpr homolog - frui
4	285.5	8.5	1072	1 A37221	neurofilament trip
5	280	8.4	3534	2 T42567	tegument protein 2
6	279	8.3	801	2 T29018	hypothetical prote
7	278	8.3	600	2 T29879	hypothetical prote
8	274.5	8.2	1299	2 T47182	hypothetical prote
9	263.5	7.9	352	2 A36128	regulatory protein
10	257.5	7.7	416	1 SKXLAG	dermal gland prote
11	252	7.5	858	2 S15762	neurofilament trip
12	250.5	7.5	352	2 G82990	neurofilament trip
13	249	7.4	1020	1 QFHUH	neurofilament trip
14	248.5	7.4	1039	2 S18199	myosin heavy chain
15	245.5	7.3	2094	2 S33224	tpr protein - huma
16	243.5	7.3	1334	2 T50568	probable multi-dom
17	243	7.3	473	2 S50755	hypothetical prote
18	242	7.2	217	2 S29309	hypothetical prote
19	242	7.2	309	2 G83013	polymyxin-like ant
20	242	7.2	915	2 S36327	clathrin assembly
21	240.5	7.2	865	2 A47282	calcium-binding pr
22	240.5	7.2	873	2 A47283	calphostin - frui
23	238.5	7.1	581	2 T22341	hypothetical prote
24	238.5	7.1	854	2 S02003	neurofilament trip
25	237	7.1	340	2 A35630	regulatory protein
26	236.5	7.1	1132	2 C75959	probable iron-sulf
27	234	7.0	901	2 A44825	phosphoprotein, sy
28	234	7.0	1110	2 T51116	NP-180 - sea lamp
29	231.5	6.9	721	2 E70766	hypothetical prote

30	231	6.9	4776	2 B95206	cell wall surface
31	230	6.9	896	2 S36326	clathrin assembly
32	229.5	6.9	1952	2 T48614	hypothetical prote
33	227.5	6.8	1367	1 S48478	glucan 1,4-alpha-g
34	226.5	6.8	1017	2 PC4035	cell-cycle-depende
35	226	6.8	822	2 A38420	antifreeze glycopr
36	226	6.8	1794	2 T38459	hypothetical diver
37	225	6.7	1931	2 A59234	slow myosin heavy
38	224.5	6.7	1736	2 T00391	hypothetical prote
39	224	6.7	441	2 T24591	hypothetical prote
40	223.5	6.7	1938	1 JX0178	myosin heavy chain
41	223	6.7	1940	2 A59287	myosin heavy chain
42	222.5	6.6	215	2 S55925	probable arabinoga
43	222.5	6.6	1133	2 T22976	hypothetical prote
44	221.5	6.6	797	1 VGBX1	glycoprotein X pre
45	221	6.6	412	2 S07537	myosin heavy chain

ALIGNMENTS

RESULT 1

QPM5H
neurofilament triplet H protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text_change 22-Jun-1999
C:Accession: J0368; A43778; S42616
R:Julien, J.P.; Cote, P.; Beaudet, L.; Sidky, M.; Flavell, D.; Grosveid, F.; Mushynski, Gene 68, 307-314, 1988
A:Title: Sequence and structure of the mouse gene coding for the largest neurofilament s
A:Reference number: J0368; MUID:89121513; PMID:3220257
A:Accession: J0368
A:Molecule type: DNA
A:Residues: 1-1087 <JUL>
A:Cross-references: GB:M23349; GB:M24496; NID:G200034; PIDN:AAA39813.1; PID:G387493
R:Shneidman, P.S.; Carden, M.J.; Lees, J.F.; Lazzarini, R.A.
Brain Res. Mol. Brain Res. 4, 217-231, 1988
A:Title: The structure of the largest murine neurofilament protein (NF-H) as revealed by
A:Reference number: A43778
A:Accession: A43778
A:Molecule type: mRNA
A:Residues: 'M', 1-132, 'QA', 134-199, 'R', 200-280, 'T', 282-491, 'G', 493-533, 'GEAKSP', 534-545,
A:Cross-references: GB:K35131; NID:G200021; PIDN:AAA39809.1; PID:G200022
R:Carden, M.J.
submitted to the EMBL Data Library, March 1994
A:Reference number: S42616
A:Accession: S42616
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-132, 'QA', 134-199, 'R', 200-280, 'T', 282-491, 'G', 493-533, 'GEAKSP', 534-545, 'R',
A:Cross-references: ENBL:Z31012; NID:G463249; PIDN:CAA83229.1; PID:G463250
C:Genetics:
A:Gene: nfh
A:Introns: 290/1; 356/3; 398/2
C:Superfamily: neurofilament triplet H protein
C:Keywords: coiled coil; cytoskeleton; heterotrimer; intermediate filament; nerve; phosph
F:1-97/Domain: amino-terminal <NTS>
F:98-408/Domain: rod #status predicted <ROD>
F:409-1087/Domain: carboxy-terminal <CTE>
F:519-886/Region: 6-residue repeats
F:520,526,532,538,544,550,556,562,568,574,580,586,592,598,604,610,616,622,628,634,640,64
73,885/Binding site: phosphate (Ser) (covalent) #status predicted
F:772/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 9.0%; Score 300; DB 1; Length 1087;
Best Local Similarity 21.6%; Pred. No. 5.1e-05;
Matches 169; Conservative 121; Mismatches 301; Indels 190; Gaps 30;

Qy 10 SYQVPADPPAIVSRAQQWVILSDENILRQELGCEYKVARLQKVETIQVRSAYEN 69
Db 53 SVSSVSASPFRCAASS-----TDSLDTLSNGPEGVVAARSEKEQLQALNDRFAG 107

```

70 LVKSSKREALEKAMRNKVLGEIRHMD-----FNDRLE-----RLETANKQL 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
108 YIDVROLEAENRS-----LEGEAAALRQQKGRAMGELYREYREMRGAVLRGAARGQL 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
114 AEKEYEGSEDTRKTIISOLFANKNSSOREKEKLEAELATARSTNEDORRHIEIRDOALS-- 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 RLEGEHLLED-----IADVROQLDEAQRREAEAAALAFAPAEAEARVELOKKAQALQ 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 -----NAQAKVVKLEELK-----KKQVYVDKVERMQQALVOLQA-ACE 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
220 EECGLRRHHOEEYGEGLGQTGGCGAAQAQAQAABARDALKCDVTSALRIIRAQLEGHAVQ 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
210 KREOLEHRLRLRLERELLESRLIQORCNCQPTNVSEVYNAALMELLREKEERTILALSA-- 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
280 SSLOSEWFVRVLDRLSEAAKVNDAMRSQOEIITYR-----RQLQARTTELEALK 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
268 ---DMTKWEQKYLEENVRHFPALDAATVAQAQDRT--TVISHSPNTSYDTALSAARIQKEE 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
332 STKESLERQSELED--RH-----QADIASYQDAIQQLDSELRNTKWMNAQLR----E 379
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
323 EEILMANKRCILDME-----GRIKTLHAQI-LEXDAMI 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
380 YQDLNLNVMAJDIIBIAYRKLLEEGECRIGFGPSPFSLTEGLPKIPISITHIKVKESEMI 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
354 KVLQQRSRKE----PSKTEQLSCMRPAKSLMSISNAGSGLLSHSLTGSPIMEEKRDDK 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
440 KVV-EKSEKETVIVEGQTEEIR-----VTEGVTEE--EDK 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
410 SWKGSGLG--ILLGDYDPAEYVPSTPSPVPSTPLLSAHSKVTGSRDOSTQTERGTENKTA 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
472 EAQOQEGEAEGEKEKEEELAAATS--PPAEAAASPEKETKSRVKEAKSPGEAKSPGE 529
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
468 AVAPISVPAPVAAAATAAAIT-ATAATITTTWAAAPVAVAAAAAPAAAAAP----- 518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
530 AKSPAFAKSPGEAKSPGCAKSPGKAPKSPAPKSPAEKSPAEKSPATVKSPGCE 589
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
519 -----SPATAATAAASVPAAGQIPAAASVASAAAAPAAAAAAVQVAPA----- 565
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
590 AKSPSEAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSP 647
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
566 APAPVPAPALVPVPAPAAQAASANOQAPTSA-----PAAVPTPTPTTPAFAVA--- 614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
648 SEAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSP 707
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
615 -----QAEVPA---SPATQPGPHRLS:PSLTCNPDKTDGPVFHSLNTERKTPQILQG 664
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
708 AEVKSPEGKSPAAVKSAPRAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSP 767
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
665 E 665
   |
768 E 768

SULT 2
4768
titferez glycopeptide AFGP polyprotein precursor [imported] - Boreogadus saida
Species: Boreogadus saida
Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
Accession: T44768
Chen, L.; DeVries, A.L.; Cheng, C.H.C.
oc. Natl. Acad. Sci. U.S.A. 94, 3817-3822, 1997
Title: Convergent evolution of antifreeze glycopeptides in Antarctic notothenioid fish
Reference numbers: Z22834; MUID:97268653; PMID:9108061
Accession: T44768
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-507 <CHE>
Cross-references: EMBL:U43200; NID:G2078482; PIDN:AAC60129.1; PID:G2078483
Genetics:
Introns: 1/3

Query Match 8.7%; Score 292; DB 2; length 507;
Best Local Similarity 35.0%; Pred. No. 4.6e-05;

```

Matches 86; Conservative 19; Mismatches 91; Indels 50; Gaps 6;

QY 433 SPVPSTLLSHSGTGRDSCSTQFERTESNKTAANAP--ISVPAPVAAATA----- 484
Db : : : : :
Db 65 TPATTAATPATAAATAAATAATAATAATAATAATAATAATAATAATAATAATA 125
QY 485 -----AAITATAATITTTWVAAPVAVAAAAAP-SPATAAAT 526
Db : : : : :
Db 126 ETPARAATPATPAATPATPAATPATPAATPATPAATPATPAATPATPAARA 185
QY 527 AAASVPAAGQIPAAASVASAAVAPSAAAAAQVAPAA-PAPVPAPALVPVPAQAQ 585
Db : : : : :
Db 186 ARAATPATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 245
QY 586 ASAP-----ACTCAPSAPAVAPTPTTPPV-----AQAEVPASP 622
Db : : : : :
Db 246 ATAPTAAATPARAARAATPATATUATAATPATPATPATPAATDATATAATPARAATPATP 305
QY 623 ATGPQP 628
Db : : : : :
Db 306 ATAATP 311

RESULT 3
T13829
Tpr homolog - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C/Accession: T13829
R.;Zimowska, G.; Aris, J.P.; Paddy, M.R.
J. Cell Sci. 110, 927-944, 1997
A;/Title: A *Drosophila* Tpr protein homolog is localized both in the extrachromosomal chan
A;/Reference number: Z17786; MUID:97296455; PMID:9152019
A;/Accession: T13829
A;/Status: Preliminary; translated from GB/EMBL/DDBJ
A;/Molecule type: mRNA
A;/Residues: 1-2346 <ZIM>
A;/Cross-references: EMBL:U91980; NID:g1923273; PID:g1923274; PIDN:AAC47506.1
C:/Genetics:
A;/Cross-references: FlyBase:FBgn0013756
A;/Map position: 2R

Query Match 8.6%; Score 286.5; DS 2; Length 2346;
Best Local Similarity 22.5%; Pred. No. 0.00041;
Matches 157; Conservative 124; Mismatches 259; Indels 157; Gaps 32;

QY 14 VPADPFALVSRAQQWEILSDENRLRGEGCYBKVA---RLQKVETEI-----Q 61
Db : : : : :
Db 1262 VGANKEEVLRIKETLNAITDSNRILREARNLTRVAELTDRISSVEKELFPQLCSNKE 1321
QY 62 RVSEAYE-NLVKSSSKRALEKAMPK--LEGEIRHMDFNEDLFERLETANKQLAEKEY 118
Db : : : : :
Db 1322 LTSKIEEINVENTSLRTBAIKWRQANALVKSNRNPEFKLQAEHLAKLLTABREL 1381
QY 119 EGSEDTRKTI-----SOLFANNKESQ---REKEKLAEALTARSTNEDORRH-I-BIRDQ 168
Db : : : : :
Db 1382 NKQSDELTVLKQPMNTPELPLMKQWILDARKQVDFTNLKQNNTROTQDIMELKNR 1441
QY 169 ALSNAQAAXVKULEELK-KQVYDVKKVMQALVOLQAACEKRQLEHLR-----TRL 222
Db : : : : :
Db 1442 LLQK-EBSILLXANEELTKDTIADTKETK-----LQLRLAKRYKDFYIGLSQGSGGTES 1496
QY 223 BRESLSRIQQRCGQCOPTNVSEYNAAALMELLREKERI-----LALEADMWKW 272
Db : : : : :
Db 1497 AAELKRVSELEEVNQ-----LRALKDEHEKITKECEVKKTBPETDSAI 1544
QY 273 EQKY---LEENVMRHFALDAATVAAQRDITTVISHSPNTSYDTALEARIQKEBSILVAN 329
Db : : : : :
Db 1545 RQEYKAKLDKLW---DLTVARTDLVNGETTFPA--GTKSSYDETI-ARLEKELOENIVAN 1598
QY 330 KCCLDMECRI----KTLHAQIIKCDAMIKVLQORSRKPSKTEQLSCVRPAKSILMSINA 385
Db : : : : :
Db 1599 K---DINORLTRENSELMRWLNQ-----LTRQLGSQSQSYKPSTS-SVA 1637

386 GSGLLSHSS-----TLTGPIMEEKDDKSKGSLGILLGGDYAEYVTPSPVPS 438
 1638 EKGNISSSPTANKWMSGSAVQOGATVPRGG-----E 1674
 439 TPLSAHSKTSRDCSTQTERGTESNKTAAVAPISVPAPVAAATAATAATATITTM 498
 1675 TPLASI-----RPIVSQNR-----TAAILPTSQPPAGSSTSTSSSSSSSTTSA 1722
 499 VAAAPVAVA-AAAAPAAAAPSPATAAATAAVSPAAA--GOIPAAASVASAAVAPSA 554
 1723 AGGSSSAVQATLVPPQQVHTTGSAALESMASSSPTSHDTPMTSPASVAVAAIPPM 1782
 555 AAAAQQVQVAPAAAPVAP-----ALVPVPAP-AAAQASAPAOQTAPT-SA 598
 1783 GASSAAESSQEAES-IOHPQNDSQLFVGGAQQVVALVSPRVEGSSSSSTSVPTATA 1841
 599 PAVAPTAPPTTAVAAQAEVAPSPATGPGPHRLSIPS 635
 1842 PSIQDGSQSQPSTSGSS--SSSSTVSSSHRHTPS 1876
 SULT 4
 7221
 urofilament triplet H protein - rat
 Species: Rattus norvegicus (Norway rat)
 Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
 Accession: A37221; A25649; A30796; A32757; B25649
 Chin, S.S.M.; Liem, R.K.H.
 Neurosci. 10, 3714-3726, 1990
 Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with vimentin
 Reference number: A37221; MUID:91038277; PMID:2320956
 Accession: A37221
 Status: preliminary; not compared with conceptual translation
 Molecule type: mRNA
 Residues: 1-1072 <CHI>
 Cross-references: GB:A301879; NID:92642597; PIDN:AAB87068.1; PID:92642598
 Robinson, P.A.; Wion, D.; Anderton, B.H.
 ES Lett. 209, 203-205, 1986
 Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).
 Reference number: A25649; MUID:87080760; PMID:2878828
 Accession: A25649
 Molecule type: mRNA
 Residues: 230-318,472-542 <ROB>
 Cross-references: GB:M37227
 Dautigny, A.; Pham-Dinh, D.; Roussel, C.; Felix, J.M.; Nusbaum, J.L.; Jolles, P.
 ochem. Biophys. Res. Commun. 154, 1099-1106, 1988
 Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ detection
 Reference number: A30796; MUID:88309090; PMID:2457365
 Accession: A30796
 Molecule type: mRNA
 Residues: 266-421, 'T', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', 727-730
 Cross-references: GB:M21964; NID:9205685; PIDN:AAA41695.1; PID:9205686
 Lieberburg, I.; Spinner, N.; Snyder, S.; Goldhaber, J.; Smulowitz, M.; O'Connell, A.C.; U.S.A. 86, 2463-2467, 1989
 Title: Cloning of a cDNA encoding the rat high molecular weight neurofilament peptide
 Reference number: A32757; MUID:89184647; PMID:2928342
 Accession: A32757
 Status: preliminary
 Molecule type: mRNA
 Residues: 553-566, 'E', 568-967, 'V', 969-997, 'GST', 1001-1022, 'E', 1024-1072 <LIE>
 Cross-references: GB:J04517; NID:9205679; PIDN:AAA41692.1; PID:9205680
 Superfamily: neurofilament triplet H protein
 Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
 Query Match 8.5%; Score 285.5; DB 1; Length 1072;
 Best Local Similarity 22.1%; Pred. No. 0.00019;
 Matches 169; Conservative 109; Mismatches 326; Indels 159; Gaps 31;
 10 SYQVPADPPAIVSRAQWVLLSDENRNLQELGCEYKAVRLQKVEIQVSEAVEN 69
 54 SVSVSASPSRFRGAASS-----TSLDTLNGFEGCVAAARSEK-QLQALNDRFAG 107

70 LVKSSKREALEKAMRNKLEGEIRRMHD-----FNRLRE-----RLETANKOL 113
 108 YIDKVRQLEAHRT-----LEGEAAALAQKGRAMGELYREVRMGAVLRLGARGHV 163
 114 AEKEYEGEDTRKTIQSOLFANKEKESQREKELKABL-ATASTNEDORRHIEIRDQALS 172
 164 RLEQHLLLED-----IAHVRLQDEARQREAEAAAALARPAQAEAAARVEL----- 212
 173 AQAKVKLESEL-----KKQVYVDKVEKMQQ--ALVQLQACEKEE----- 212
 213 -OKKAQALQEGCVLRHHEVEGELGQIQGGAAQAQAQAARDALKCDVTSAUREIR 271
 213 -----QLEHLRLTRLERELSLRIOQRQNCOPTNVSEYNAALMELLREKEE 260
 272 AQLBHTVQSTLQSEWFRVRLDLSEAAKVNWDAMKSAQEEITEYR-----RQLQA 323
 261 RILALEA-----DMTWKQKYLEENVMHFPALDAATVAQORDTTVISHSPNTSYDALE 315
 324 RTTELEAKSTKESLERQSELED-----RH-----QVDMASYQD--AIQQLDNEIRNTKWE 373
 316 ARIQKEEEELMANKRCLDME-----GRKTLHAQI- 346
 374 MAQQLREVQDLLNVQWALDIEIAAYRKLLEGECRIGPGSPFSLTEGLPKIPSMSTHIX 433
 347 IEKDAMIKVLOORSKE-----PSKTEQLSCMRPA-----KSLMSISNAGSGLLSHSSTLT 397
 434 VKSEKIKVW-EKSEKETVIVEQTEIEQVTEETEEDEKBAQGESEEEAEGBEGEAATT 492
 398 GSPIMERKDDKSKGSLGILLGGDYAEYVTPSPVPPSTPLLSAHSKTSRDCSTQT 457
 493 SPFAEAAKSPKETSQV-----KBAKSPAKSPAEAKSP--AEAKSPAEVKSAPAE 543
 458 ERGTESNKTAAVA--PISVPAPVAAAATAAATATATATITTTWVAAPV-AVAAAAPAA 514
 544 VKSPAEAKSPAEAKSPAEVKSAPATVKSPAEAKSPAEAKSPAEVKSAPATVKSPAE 603
 515 AAP-----SPATAATAAAVSPAAA--GOIPAAASVASAAVAPSAVAAVAPAPAP 567
 604 AKSPAEVKSAPAEAKSPAEVKSAPVSPGAEKSPAEAKSPAEVKSAPATVKSPVEA--KSP 661
 568 APVPAPALVPVAPAAA--QASAPAOQTAPSA--PAVAPTAPPTTTPVAAQAEVPA-SP 622
 662 AEVKSAPVTVKSAPAEKSPVVEVKSAPVKSAPAEKSPAEKSPAEKSPVVKSPAEAKSP 721
 623 ATGCGPERLSLTCNPKDTGDFVHSNTLEKTIQILQGE 665
 722 AEAKPPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPVVEVKSPE 764
 RESULT 5
 T42567
 tegument protein 24 - equine herpesvirus 4 (strain NS80567)
 C:Species: equine herpesvirus 4
 A:Variety: strain NS80567
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T42567
 R:Telford, B.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
 J. Gen. Virol. 79, 1197-1203, 1998
 A:Title: The DNA sequence of equine herpesvirus-4
 A:Reference number: Z22173; MUID:98264497; PMID:9603335
 A:Accession: T42567
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3534 <TEL>
 A:Cross-references: EMBL:AF030027; NID:92605950; PIDN:AACS9539.1; PID:92605967
 A:Experimental source: strain NS80567
 C:Genetics:
 A:Note: 24
 C:Superfamily: varicella-zoster virus gene 22 protein
 Query Match 8.4%; Score 280; DB 2; Length 3534;
 Best Local Similarity 34.4%; Pred. No. 0.0012;

Query Match	7.5%; Score 250.5; DB 2; Length 352;
Best Local Similarity	25.4%; Pred. No. 0.0014;
Matches 107; Conservative 42; Mismatches 170; Indels 103; Gaps 13;	
QY	212 EOLEHRLATRLERLESIRIQRGNGCOPINVSYNAAALMELLEKEBERILALEADMTK 271 :: :: DB 16 QQLSHSL-----VEHLEGACK-----QAIVDSFKLAKLEKQRGK 50
QY	272 WEQKYLEENVNRHFALDAAATAAQRDDTTVISHSPNTSYDTALEARIQKEEETILMANVR 331 :: :: DB 51 AQEKLLHKARTKLQDAAGAGTKAQAKARETIS-DLEBALDT-LKAR-QADTRTYIVGLKR 107
QY	332 CLUDMSGRIKTLHAQIIEKDAMIKVIQORSRKBSKTEQLSCWRPAPKSLMSINSAGSGILLS 391 :: :: DB 108 --DVQESLKAGGVKVKAEAGKALBSRKAKPATPKAAPAAAAPKA-----150
QY	392 HSTLTGTGPIMEEKRDDKSWKSGSILGGDYRAEVVPSTPPSPVPPTLLSAHSKTGSR 451 :: :: DB 151 -----VKTVAAKPAKPAKPAKPAKPAKPAKPA-176
QY	452 DCGSTQTERTGSNKTAAVAPISVPAPVAAAAATAAAITA-TAATITTTVVAAAPVAVAAAA 510 DB 177 -----KTAAPAKPAKPTAKPAKPAKPAKPAKTAAPAKPAKPAKPAKPAKPAK 223
QY	511 AAAAAAASPDATAATAAVNSPAAAGOIIPAAASVTSAAAVAPSAAAAAAVQP-----AA 566 DB 224 PAAKTAAPAKPAKPAKPAKVAPTAK---PAAKTAAPAKPAKPAKPAKPAKPAKPAKPAK 280

510 ANAKSSLE--KAKHRLQNEQEDWMDLEKANSAAASLDKKQSGPKINDWKQYEE 566
 366 KTEQSCMRPAKS---LMSISNAGSGLLSHSSTLTGSPIMEEKRDDKS 410
 567 QAELEASQKASLSLTELFLKNAVEETLDHLETL-----KLENKN 607

protein - human
 alternate names: kinase-related transforming protein (tpr-met); protein with promoter
 species: Homo sapiens (man)
 date: 22-Nov-1993 #sequence revision 26-May-1995 #text_change 11-Jan-2002
 accession: S33124; S23740; S00928; G01185
 itchell, P.J.; Cooper, C.S.
 cogene 7, 2329-2333, 1992
 title: The human tpr gene encodes a protein of 2094 amino acids that has extensive coiled-coil structure
 reference number: S33124; MUID:93064711; PMID:1437155
 accession: S33124
 tatus: preliminary; nucleic acid sequence not shown; translation not shown
 clecule type: mRNA
 esidues: 1-2094 <MIT>
 ross-references: EMBL:X66397; NID:9633225
 ste: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
 itchell, P.J.; Cooper, C.S.
 cogene 7, 383-388, 1992
 title: Nucleotide sequence analysis of human tpr cDNA clones.
 reference number: S23740; MUID:92195670; PMID:1549355
 accession: S23740
 tatus: preliminary
 clecule type: mRNA
 esidues: 1-725, 'L' <M12>
 ross-references: EMBL:X63105; NID:g37257; PIDN:CAA44819.1; PID:g37258
 ing, H.W.S.; Tempest, P.R.; Merrifield, K.R.; Rance, A.J.
 cogene 2, 617-619, 1988
 title: Tpr homologues activate met and raf.
 reference number: S00928; MUID:88262257; PMID:3387099
 accession: S00928
 clecule type: mRNA
 ssidues: 1-31, 'R', 33-142 <KIN>
 ross-references: EMBL:Y00672; NID:g37255; PIDN:CAA68681.1; PID:g37256
 reco, A.
 nitted to the EMBL Data Library, December 1995
 reference number: H00592
 accession: G01185
 atus: translated from GB/EMBL/DBJ
 clecule type: DNA
 ssidues: 144-228 <GRE>
 ross-references: EMBL:X94208; NID:g1296797; PIDN:CAA63904.1; PID:g1296798
 netics:
 ne: GDB:TPR
 ross-references: GDB:128821; OMIM:189940
 p position: 1q25-1q25
 tions: 177/3

ery Match 7.3%; Score 245.5; DB 2; Length 2094;
 st Local Similarity 20.3%; Pred. No. 0.016;
 tches 153; Conservative 127; Mismatches 274; Indels 201; Gaps 31;

24 RAQQWVEILSDENRLRQEGCEKVV---ARLQKVETIORVSEAVENLVKSSK--- 76
 1223 RYRQVELLELELEQLESLNAEKQVVTAKTQAQHELMKKTEIM-NVMTNKLRE 1281

77 -REALEKAMR-----NKLGEIRRMDFNRDLER--LETANKQLAEKYE----- 119
 1282 EKERLEQDLQMQAKVRKLELDILPQEAANAESEKGMQLQAEKLEEDVKWKARNQH 1341

120 -----GSEDTKNTISQLPAKKEKESREKE---KLEAEIATARSTNEDQRHIEIRDQ 168
 1342 LVSQQKDPDTEYRKLEKEVHTKRIQQLTEIGRLKAEIARNA----- 1387

169 ALSNAQAKVVKLEELKKKKVYVDKVEK-MQQAIVQLQAAACEKREBLEHRLRLRLERELE 227

Db 1388 SLTNNQMLIOSLKEDLNKVRTEKETIQKDLDAKIIDIQEKVKTIQVK-KIGRRYKTOYE 1446
 QY 228 SLRIO-----ORQGNCOPTNVG-----EYNAALMELLRE 257
 Db 1447 ELKAOQKWMETSQAQSGSDHQBQHVSVQEQELKETLQAETKSKSLSQVENLQKTLSE 1506
 QY 258 KB-----ERILALGADMTKWEQKYLENVNMRHPALDAAATVAAQORD-TTVVISHSPNTS 309
 Db 1507 KETEARNLOBCTVQLQSELRLQD-LQDRITTOBQLROQITEKEBKTKAIVAAKSKIA 1565
 QY 310 YDTALEARIQKEEIEILMAN-----KRCIDM-----EGRIKTLHAQIIEKDAMIK 354
 Db 1366 HLAGVKDOLTKENBELKQNGALDQOKDELVDRIITALKSOYEGRISRLERELREHQB--R 1623
 QY 355 VLQOR-----SRKGPSTKTEQLSCMRPAKSIAMISINAGSGLLSHSSTLTGSPIMEEKRDD 408
 Db 1624 HLEQRDBEPQPSNKPVEQQRQI-----TLKTTFA----- 1652
 QY 409 KSMKGSGLILGCDYRAEYVPSTPSPVPPSTPLLSAHSKTSGRDCSTQTERGFSNKTTAA 468
 Db 1653 ---SGERGIASTSD-----PPTANIKP--TPVSTPSKV-----TAAAMAGNKSTPRAS 1696
 QY 469 VAPISVPAPVAAAATAATATAATATTWAAAAPVAVAAAAAPAAAAAPSPA-TAAATA 527
 Db 1697 IRPMVTETATVNTPT-----TPTATVMTTQVSEOEAMQSEGVEHVVPVFGSTGSGVRSTS 1752
 QY 528 AAVSPAAAGQIPAAASVASAAAAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 587
 Db 1753 PNYQPSISQPI-----LTVOOQTOATFVQ-----PTQOSHPIEPANQELSSNIV 1798
 QY 588 APAQT---QAPTSAPAVATPATPTTPAV-----AQAEVPASPATGPGPHR 630
 Db 1799 EYVQSSPVPERESTATVFGTVSATPSSSLPKTRBEEDSTIENSQVSDDTVEMPLEPKK 1858
 QY 631 LSTPSILTCNPKDKTDGVPFHSNLTLEKTPIQILQOE 665
 Db 1859 LK---SVTFVGTETEEVMAEESDGEVETQVYNQD 1869

Search completed: June 16, 2004, 19:15:39
 Job time : 31.2311 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: June 16, 2004, 19:12:42 ; Search time 16.5037 Seconds
(without alignments)
2129.669 Million cell updates/sec

le: US-09-332-063-2

fect score: 3347
pience: 1 MPRAQPSASVQVPADPFA.....KTPIQILQEPDAEMVEYLI 675

ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

al number of hits satisfying chosen parameters: 141681

imum DB seq length: 0
imum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	300	9.0	1087	1 NFH_MOUSE	P19246 mus musculus
2	292.5	8.7	518	1 TPM4_DROME	P49455 drosophila
3	257.5	7.7	439	1 XP2_XENLA	P17437 xenopus lae
4	252	7.5	857	1 NFM_CHICK	P16053 gallus gall
5	250.5	7.5	352	1 ALGE_PSEAE	P15276 pseudomonas
6	246	7.3	1026	1 NFH_HUMAN	P12036 homo sapien
7	245.5	7.3	2349	1 TPR_HUMAN	P12270 homo sapien
8	242.5	7.2	1102	1 MYSC_CHICK	P29616 gallus gall
9	242	7.2	915	1 A180_RAT	Q05140 rattus norv
10	241.5	7.2	831	1 NFM_BOVIN	P16884 rattus norv
11	241.5	7.2	865	1 CPN_DROME	Q02910 drosophila
12	240.5	7.2	865	1 BECI_YEAST	P47068 saccharomyc
13	239.5	7.1	1157	1 BECI_YEAST	Q8b166 mus musculus
14	239	7.1	790	1 BEAI_MOUSE	Q61548 mus musculus
15	234	7.0	901	1 A180_MOUSE	Q10690 mycobacteri
16	231.5	6.9	721	1 YK82_MYCTU	Q9wve9 rattus norv
17	230.5	6.9	1217	1 ITN1_RAT	Q90339 cyprinus ca
18	229.5	6.9	1935	1 MYSS_CYPCA	Q03173 mus musculus
19	229	6.8	802	1 ENAH_MOUSE	P08640 saccharomyc
20	227.5	6.8	1367	1 AMYH_YEAST	P49454 homo sapien
21	226.5	6.8	3210	1 CNP_HUMAN	P24856 notothenia
22	226	6.8	790	1 ANP_NOTTO	Q10172 schizosacch
23	226	6.8	1794	1 YDC9_SCHPO	Q06041 homo sapien
24	224.5	6.7	907	1 A180_HUMAN	P13338 gallus gall
25	223.5	6.7	1938	1 MYSS_CHICK	Q9y313 homo sapien
26	223	6.7	622	1 BPPI_HUMAN	P28968 equine herp
27	221.5	6.6	797	1 VGLX_HSVB	P14105 gallus gall
28	221	6.6	1959	1 MYH9_MOUSE	Q920r4 mus musculus
29	218.5	6.5	1714	1 ITN1_CHICK	Q15811 homo sapien
30	218	6.5	1721	1 ITN1_HUMAN	Q99105 oryctolagus
31	217.5	6.5	501	1 MYSU_RABIT	Q9j1c0 rattus norv
32	217.5	6.5	1976	1 MYHA_RAT	P24733 acquiripeten
33	217	6.5	1938	1 MYSS_REQIR	

ALIGNMENTS

RESULT 1

ID	NFH_MOUSE	STANDARD	PRT	1087 AA
AC	P19246; Q61359;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Neurofilament triplet H protein (200 kDa neurofilament protein)			
DE	(Neurofilament heavy polypeptide) (NF-H).			
GN	NFPH OR NFH.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89121513; PubMed=3220257;			
RA	Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,			
RA	Mushynski W.;			
RT	"Sequence and structure of the mouse gene coding for the largest			
RT	neurofilament subunit.";			
RL	Gene 68:307-314(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=69089138; PubMed=3145094;			
RA	Shneidman P.S., Carden M.J., Lees J.P., Lazzarini R.A.;			
RT	"The structure of the largest murine neurofilament protein (NF-H) as			
RT	revealed by cDNA and genomic sequences.";			
RL	Brain Res. 464:217-231(1988).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Swiss Webster; TISSUE=Brain;			
RA	Carden M.J.;			
RL	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: Neurofilaments usually contain three intermediate			
CC	filament proteins: L, M, and H which are involved in the			
CC	maintenance of neuronal caliber. NF-H has an important function in			
CC	mature axons that is not subserved by the two smaller NF proteins.			
CC	!- PTM: There are a number of repeats of the tripeptide K-S-P. NFH is			
CC	phosphorylated on a number of the serines in this motif. It is			
CC	thought that phosphorylation of NFH results in the formation of			
CC	interfilament cross bridges that are important in the maintenance			
CC	of axonal caliber.			
CC	!- PTM: Phosphorylation seems to play a major role in the functioning			
CC	of the larger neurofilament polypeptides (NF-M and NF-H). The			
CC	levels of phosphorylation being altered developmentally and			
CC	coincident with a change in the neurofilament function.			
CC	!- SIMILARITY: Belongs to the intermediate filament family.			
CC	!- CAUTION: Ref.2 sequence differs from that shown in positions 534			
CC	to 716 and is shorter due to frameshifts.			

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M24496; AAA39813.1; JOINED.
EMBL; M23349; AAA39813.1; JOINED.
EMBL; M24494; AAA39813.1; JOINED.
EMBL; M24495; AAA39813.1; JOINED.
EMBL; M35131; AAA39809.1; ALT_FRAME.
EMBL; M31012; CAA83229.1; --
PIR; JTO368; QFMNSH.
MGD; MGI:97109; Nefh.
InterPro; IPR001664; IF.
Pfam; PF00038; filament; 1.
PROSITE; PS00226; IF; 1.
Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
DOMAIN 1 97
HEAD.
DOMAIN 98 408
TAIL.
DOMAIN 409 1087
GLU-RICH (ACIDIC).
DOMAIN 436 517
50 X 6 AA TANDEN REPEATS OF K-S-P-A-E-A.
DOMAIN 519 886
GLU/LYS-RICH.
DOMAIN 887 1087
COIL 1A.
DOMAIN 98 129
LINKER 1.
DOMAIN 130 141
LINKER 1B.
DOMAIN 142 239
LINKER 12.
DOMAIN 240 261
COIL 2A.
DOMAIN 262 283
LINKER 2.
DOMAIN 284 287
COIL 2B.
DOMAIN 288 408
K -> QA (IN REF. 2 AND 3).
A -> AR (IN REF. 2 AND 3).
S -> T (IN REF. 2 AND 3).
L -> G (IN REF. 2 AND 3).
P -> PREAKSP (IN REF. 3).
MISSING (IN REF. 3).
G -> A (IN REF. 3).
V -> M (IN REF. 2 AND 3).
T -> N (IN REF. 2 AND 3).
SEQUENCE 1087 AA; 116612 MW; 57BAC76A38EDICB9 CRC64;
Query Match 9.0%; Score 300; DB 1; Length 1087;
Best Local Similarity 21.6%; Pred. No. 5.1e-05;
Matches 169; Conservative 121; Mismatches 301; Indels 190; Gaps 30;
10 SYQVPDPDFPAIVSRAQQMVEILSDENRNLQELGECYKVARLQVETIQRVSAYEN 69
53 SVSSVSASPSRFGAASS-----TDSLDTLSNGPCCVAAVAARSEKEQLQALNDRFAG 107
70 LVKSSKREALEKAMRNKLEGEIRMHED-----FNEDLRE-----RLETANKOL 113
108 YIDKVRQLEAHNR-----LEGEAAALQOQKRAAMGELYERVMRGAVLRUGAARGQL 163
114 AEKSYEGSEDTRKTSIQLFANKESQREKEKLEAELATARSTNEDQRRHTEIRDQALS-- 171
164 RLEQEHLLD-----IAHVQRQLDEEARQREAEAEARALAPAEAEAAARVELQKKAALQ 219
172 -----NAQAVKLEBEUK-----KQVTVDKVKNQQAALVQLQA-ACE 209
220 BECYLRRHHQHEVGEGLGQGGCAQAQAQAQAPDALACDVTSALREIRAQLEHGVQ 279
210 KEOLEHLRLTRLELESRLIQRCQNCQPTNVSEYNAALMELLREKEERILALEA-- 267
280 SLSQSEFWFRVRLDLSEAAKVNVDAMSQAQSEITEYR-----RQLQARITTEALK 331
268 ---DMTKWEQKYLENVNRHFALDAATVAQAORT--TVISHPNTSYDPALEARTQKEE 322
332 STKESLERQSELED---RH-----QADIASYQDAIQQLDSELANTKWMAAQLR-----E 379
323 EELIMANKSCLDME-----GRKTLHAQI-IEKDAMI 353
380 YQDLNVQALDIEIAVNRKLEGECEICIGPSPFSLTGLPKIPISITHIKVSEMI 439
354 KVLQORSRKE-----PSKTEQLSCMRPAKSLMSISNAGSLSHSSTLTGSPIMEEKDDK 409

440 KVV-EKSEKETVIVEGQTEIR-----VTEGVTE--EDK 471
410 SWKSLG--ILLGDYRAEYVPSTPSPSTPLLSAHSKTSGRDCSTOTERTESNKTA 467
472 ENQOGECEEAEEGEKEEELAAATS--PPAEAAAPKEKETSRVKEAKSPGEAKSPGE 529
468 AVAPISVPFVAAAATAAAT-ATAATITTTTAAAPVAVAAAAAFAAAAP----- 518
530 AKSPAEAKSPGEAKSPGEAKSPGEAKSPGEAKSPGEAKSPGEAKSPGEAKSPGE 589
519 -----SPATAATAAAVSPAAAGQIPAAASVASAAAVAPSAASAAAVQVAPA----- 565
590 AKSPSEAKSPAEAKSPAEAKSPA-EA-KSPA-EAK-SPA-EAKSPA-EAKSPA-EAKSPA 647
566 APAPVPALVPAPAPAAQAASAPAQTAQTS-----RAVATPAPTTPAVA--- 614
548 SEAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA 707
615 -----QAEVPA-----SPATGPQPHRLSLPSLTCNDPKTDGPPVHSHNTLERTPIQLGQ 664
708 AEVKSPEGKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA-EAKSPA 767
665 E 665
768 E 768
RESULT 2
ITEM4 DROME
ID TPW4 DROME STANDARD; PRT; 518 AA.
AC P49455; P49456; Q24425; Q24426;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tropomyosin 1, isoforms 33/34 (tropomyosin II).
GN TMI OR TMIIL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC TISSUE=Embryo, and Pupae;
RX MEDLINE=89127197; PubMed=2851721;
RA Hanke P.D., Storti R.V.;
RT "The Drosophila melanogaster tropomyosin II gene produces multiple proteins by use of alternative tissue-specific promoters and alternative splicing."
RT Mol. Cell. Biol. 8:3591-3602 (1988).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC STRAIN=Oregon-R; TISSUE=Pupae;
RX MEDLINE=87064486; PubMed=3097506;
RA Karlik C.C., Fytberg E.A.;
RT "Two Drosophila melanogaster tropomyosin genes: structural and functional aspects".
RT Mol. Cell. Biol. 6:1965-1973 (1986).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Comment=Additional isoforms seem to exist;
CC Name=33; Synonyms=9C;
CC IsoId=P49455-1; Sequence=Displayed;
CC Name=Muscle; Synonyms=9D;
CC IsoId=P06754-1; Sequence=External;
CC Name=Non-muscle; Synonyms=Cytoskeletal;
CC IsoId=P06754-2; Sequence=External;
CC Name=9A;
CC IsoId=P06754-3; Sequence=External;
CC Name=34; Synonyms=9B;
CC IsoId=P49455-2; Sequence=VSP_006623, VSP_006624, VSP_006625;
CC -!- TISSUE SPECIFICITY: Both isoforms are only expressed in indirect flight muscles.

epidermis. May also be involved in growth of regenerating glands and in protection of the skin from the external environment.

-!- SUBCELLULAR LOCATION: Secreted.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Name=1; Synonyms=APEG;
 IsoId=P17437-1; Sequence=Displayed;
 Name=2; Synonyms=XP2;
 IsoId=P17437-2; Sequence=VSP_004652;
 -!- TISSUE SPECIFICITY: Skin.
 -!- SIMILARITY: Contains 2 P-type (trefoil) domains.
 -!- CAUTION: Ref.2 sequence differs from that shown from position 392 onward and is shorter (418 AA) due to a frameshift.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M90095; AAA50001.1; --
 EMBL; X51394; CAA35759.1; ALT_FRAME.
 PIR; A37331; A37331.
 PIR; S07498; SKXLAG.
 HSSP; P04155; 1PS2.
 InterPro; IPR000519; P_trefoil.
 Pfam; PF00088; trefoil_2.
 PRINTS; P00680; PTFREFOIL.
 SMART; SM00018; PD; 2.
 PROSITE; PS00025; P_TREFOIL; 2.
 Signal; Growth factor; Alternative splicing; Repeat;
 Pyroglutamate carboxylic acid.
 SIGNAL 1 22 POTENTIAL.
 CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.
 MOD RES 23 23 PYROLIDONE CARBOXYLIC ACID (PROBABLE).
 DCMAIN 26 343 33 X REPEATS OF G-[GE]-[AF] (2,4)-A-E.
 DOMAIN 350 391 P-TYPE 1.
 DOMAIN 397 438 BY SIMILARITY.
 DISULFID 351 376 BY SIMILARITY.
 DISULFID 361 377 BY SIMILARITY.
 DISULFID 371 388 BY SIMILARITY.
 DISULFID 398 424 BY SIMILARITY.
 DISULFID 408 423 BY SIMILARITY.
 DISULFID 418 435 BY SIMILARITY.
 VARSPLIC 26 343 Missing (in isoform 2).
 CONFLICT 3 3 H -> S (IN REF. 2).
 CONFLICT 18 18 C -> W (IN REF. 2).
 SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;

Query Match 7.78; Score 257.5; DB 1; Length 439;
 Best Local Similarity 34.2; Pred. No. 0.00088;
 Matches 83; Conservative 16; Mismatches 113; Indels 31; Gaps 8;
 432 PVPPTPLLSAHSKGTGRDCTQTERGFTSNKTAAPISVPAPVAAAATAAATA 491
 119 PAPAPAGEAPAPAPAGEAPAPAPAE-GEAPAPAGEAPAPAPAEVEAPAPAGEAP 177
 492 ATITTTTVAAPVAV-----AAAAPAAAAPSPATAATAAASVPA-AACGIPA----- 540
 178 APAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGE 237
 541 --AASVAAAAPSAAPAAAQVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 595
 238 APAPAPAGEAPAPAPAGEAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAGEAP 297
 596 TSAPAVAPTPTP-----TPAVAAQEVPA-----SPATGPGPHLSLPTCNPD--- 641
 298 APAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAPAPAPAPAPAPAPAPAPAP 357
 642 KTD 644

Db 358 RTD 360
 RESULT 4
 NFM CHICK
 ID NFM CHICK STANDARD; PRT; 857 AA.
 AC P16053;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurofilament triplet M protein (160 kDa neurofilament protein)
 DE (Neurofilament medium polypeptide) (NF-M).
 GN NFM.
 OS Gallus Gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 OX NCBI_TaxID=9031;
 [1]
 RF SEQUENCE FROM N.A.
 RA MEDLINE=90174973; PubMed=2106668;
 RA Zopf D., Dineva B., Betz H., Gundelfinger E.D.;
 RT "Isolation of the chicken middle-molecular weight neurofilament
 RT (NF-M) gene and characterization of its promoter.";
 RL Nucleic Acids Res. 18:521-529(1990).
 RN [2]
 RF SEQUENCE OF 259-857 FROM N.A.
 RA MEDLINE=88112814; PubMed=3123320;
 RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D.; Betz H.;
 RT "Identification of gene products expressed in the developing chick
 RT visual system: characterization of a middle-molecular-weight
 RT neurofilament cDNA.";
 RL Genes Dev. 1:699-708(1987).
 CC -!- FUNCTION: Neurofilaments usually contain three intermediate
 CC filament proteins: L, M, and H which are involved in the
 CC maintenance of neuronal caliber.
 CC -!- PTM: There are a number of repeats of the tripeptide K-S-P, NFM is
 CC phosphorylated on a number of the serines in this motif. It is
 CC thought that phosphorylation of NFM results in the formation of
 CC interfilament cross bridges that are important in the maintenance
 CC of axonal caliber.
 CC -!- PTM: Phosphorylation seems to play a major role in the functioning
 CC of the larger neurofilament polypeptides (NF-M and NF-H), the
 CC levels of phosphorylation being altered developmentally and
 CC coincident with a change in the neurofilament function.
 CC -!- SIMILARITY: Belongs to the intermediate filament family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; X17102; CAA34958.1; --
 EMBL; X05558; CAA29073.1; --
 PIR; S15762; S15762.
 InterPro; IPR006821; Filament_head.
 InterPro; IPR001664; IF.
 InterPro; IPR002957; Keratin_I.
 Pfam; PF00038; Filament; 1.
 Pfam; PF04732; filament head; 1.
 PRINTS; PR01248; TYPE1KERATIN.
 PROSITE; PS00226; IF; 1.
 Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 Glycoprotein.
 INIT_MET 0 0
 FT DOMAIN 1 98 HEAD.
 FT DOMAIN 99 406 ROD.
 FT DOMAIN 407 857 TAIL.
 FT DOMAIN 99 130 COIL 1A.

DOMAIN 131 143 LINKER 1.
 DOMAIN 144 242 COIL 1B.
 DOMAIN 243 259 LINKER 12.
 DOMAIN 260 281 COIL 2A.
 DOMAIN 282 285 LINKER 2.
 DOMAIN 286 406 COIL 2B.
 CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
 CARBOHYD 426 426 O-LINKED (GLCNAC) (BY SIMILARITY).
 CONFLICT 546 546 G -> R (IN REF. 2).
 SEQUENCE 857 AA; 95704 MW; 4B2E0FC6AC64778B CRC64;

Query Match 7.5%; Score 252; DB 1; Length 857;
 Best Local Similarity 19.7%; Pred No. 0.0028;
 Matches 140; Conservative 104; Mismatches 248; Indels 218; Gaps 22

```

27 QMVEITLSDENRNLRELGGCGVEKVARIQ-----KVETIQRVSEAYENLVKSSKKRALEK 82
   : : | | | | | : : : | | | | : : | | : : | |
139 QLGDYAEQELRLRGALGVSHKAQIQLDSEHTIEDIQLRRFRFDEARLRDEATEATIA 198
   : : | | | | | : : : | | | | : : | | : : | |
83 AVENKLGEIEIRMDIFNRLDERLETANKOLAEKYEGSBDTKTISQLPFKKESOREK 142
   : : | | | | | : : : : : : : : : : : : : : :
199 ALRKEME-EASLM---RAELDOKVQSLODEVAFURGNHBESEVALLAQAASHATVER-K 253
   : : | | | | | : : : : : : : : : : : : : : :
143 EKLAEALATA-----RSTNEDORRH-----IEIRDQALSNAQAKVKV 179
   : : | | | | | : : : | | | | : : | | : : | |
254 DYLTGDTLTALKEIRAQLCEQSDHNMQABEMFKRYAKLTEAAEQNKGAIRSKEEIAE 313
   : : | | | | | : : : | | | | : : | | : : | |
180 LSELKKQGVYDKYKVKQQALVOLQACEKR-----SOLEHLRL-TRLE---- 223
   : : | | | | | : : : | | | | : : | | : : | |
314 YRQIQSKSIIELESVRGTKESLERQLSDIERHNNDLTTYYODTHOLENELRGTKWEMAR 373
   : : | | | | | : : : | | | | : : | | : : | |
224 --RELIESIKRIQRQNCQPTNVIS--BYNMAALMELLREKEEIIALEADMT----- 270
   : : | | | | | : : : | | | | : : | | : : | |
374 HLREYODL-----LNWKWALDIENAYEKLEGSETSPFSFGSIITPIFTHRQ 422
   : : | | | | | : : : | | | | : : | | : : | |
271 -----KWOKYLEENV-----MRHPALDAAAIVAAQRDTPVIS 303
   : : | | | | | : : : | | | | : : | | : : | |
423 PSVTIASTKIQTKEIPPKLVQHKFVEEIIETEKTVEDKSEMEDALSATABE----- 475
   : : | | | | | : : : | | | | : : | | : : | |
304 HSPNTSYDTALARIQEEEBILANVRCLDMEGRIKTLHAQIIIEKDAMIKVLQORSKE 363
   : : | | | | | : : : | | | | : : | | : : | |
476 -----MAKQAQEEGQBEEKABE-----EAVEEENVEKVAEQAAEE 511
   : : | | | | | : : : | | | | : : | | : : | |
364 PKSTEQLSCMRPAKSMGISNAGSGLLSHSSTLTGSPIMBEKRDKSXGSLGILLGGDY 423
   : : | | | | | : : : | | | | : : | | : : | |
512 BEKEEEEAEEREAAKSDAAEBSGSKKEEIEBEKSGEBAEERBAKGAEEAGA----- 565
   : : | | | | | : : : | | | | : : | | : : | |
424 RAELYVPSTPSVPVPPSTPLSAHSKTGSRDCSTQTERGTSENKTAAPAIVTSVPAPVAAAAT 483
   : : | | | | | : : : | | | | : : | | : : | |
566 KVEKVKSPKSPKSPKSP-----PSPATAAATAAASPAAAAA----- 582
   : : | | | | | : : : | | | | : : | | : : | |
484 AAAITATAATITTTVVAAAPVAVAAAAAAPAAAAA-----PSPATAAATAAASPAAAAAQGP 539
   : : | | | | | : : : | | | | : : | | : : | |
583 -----PKSPYTEQAKAVQKAAEVGKDQAEKAAEKAAKEKAAASPEKP 626
   : : | | | | | : : : | | | | : : | | : : | |
540 AASVASAANAAPSAASAAAADVMPAPAP--APVPAPALVPVPAPAAAAQASAQAQCPTS 597
   : : | | | | | : : : | | | | : : | | : : | |
627 ATPKVTSP-----XPATPEKPTPEKAITPEKVKRSEPKTTEPKVUSP-E 671
   : : | | | | | : : : | | | | : : | | : : | |
598 APAVAPTAPTPTPAVAQEAEPASPGATGCPGPHRLSIPSILTCNFDKTDGPV 647
   : : | | | | | : : : | | | | : : | | : : | |
672 KPASEKPTPEKP--ASPEKATPEKPTPEKATPEKPRSEPEKSSPL 719
   : : | | | | | : : : | | | | : : | | : : | |

```

MULT 5
 iP_PSEAE STANDARD; PROT; 352 AA.
 P15276; Q9HTU1;
 01-APR-1990 (Rel. 14, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Transcriptional regulatory protein algp (Alginate regulatory protein algr3).

ALGP OR ALGR3 OR PA5253.
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=287;
[1]_TaxID=287;
SEQUENCE FROM N.A.
RN
RP STRAIN=8822;
RC MEDLINE=90108714; PubMed=2514124;
RX Kato J., Chu L., Kitano K., Devault J.D., Kimbara K.,
RA Chakraborty A.M., Misra T.K.;
RT "Nucleotide sequence of a regulatory region controlling alginate
RN synthesis in Pseudomonas aeruginosa: characterization of the algr2
RL Gene."; ;
RL Gene 84:31-38(1989).
[2]
SEQUENCE FROM N.A.
RN
RP STRAIN=8882;
RC MEDLINE=9022135; PubMed=2109318;
RX Kato J., Misra T.K., Chakraborty A.M.;
RA "Algr3, a protein resembling eukaryotic histone H1, regulates
RN alginate synthesis in Pseudomonas aeruginosa."; ;
RL Proc. Natl. Acad. Sci. U.S.A. 87:2887-2891(1990).
[3]
SEQUENCE FROM N.A.
RN
RP STRAIN=8830;
RC MEDLINE=91008921; PubMed=1698761;
RX Deretic V., Konyscsni W.M.;
RA "A prokaryotic regulatory factor with a histone H1-like
RN carboxy-terminal domain: clonal variation of repeats within algp, a
RL gene involved in regulation of mucoidy in Pseudomonas aeruginosa."; ;
J. Bacteriol. 172:5544-5554(1990).
[4]
SEQUENCE FROM N.A.
RN
RP STRAIN=PAO / PAO2003;
RC MEDLINE=90236911; PubMed=2110144;
RX Konyscsni W.M., Deretic V.;
RA "DNA sequence and expression analysis of algp and algl, components of
RN the multicene system transcriptionally regulating mucoidy in
RL Pseudomonas aeruginosa: algp contains multiple direct repeats."; ;
J. Bacteriol. 172:2511-2520(1990).
[5]
SEQUENCE FROM N.A.
RN
RP STRAIN=ATCC 15692 / PA01;
RC MEDLINE=20437337; PubMed=10984043;
RX Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman P.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiser J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;
RN "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen.";
RL Nature 406:959-964(2000).
-!- FUNCTION: THE PROMOTER FOR A CRITICAL ALGINATE BIOSYNTHETIC
GENE, ALGD, ENCODING GDP-MANNOSE DEHYDROGENASE, IS ACTIVATED ONLY
UNDER CONDITIONS REMINISCENT OF THE CYSTIC FIBROSIS LUNG (I.E.,
UNDER HIGH OSMOLARITY), AND AT LEAST TWO REGULATORY GENES, ALGP
AND ALGL, HAVE BEEN IMPLICATED IN THIS ACTIVATION PROCESS.
-!- DOMAIN: THE CARBOXY-TERMINAL BINDS TO DNA. IT IS UNKNOWN WHETHER
BINDING IS SPECIFIC OR NON-SPECIFIC.
-!- SIMILARITY: TO EUKARYOTIC HISTONES H1.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

EMBL; M30145; AAA25724.1; -;
DR EMBL; M35259; AAA25705.1; -;

Daigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L., Jolles P.;
"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ detection";
Biochem. Biophys. Res. Commun. 154:1099-1106(1988).

[3]
SEQUENCE OF 1-89 AND 243-313 FROM N.A.
MEDLINE=87080760; PubMed=2878828;
Robinson P.A., Wion D., Anderson B.H.;
"Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H)";
FEBS Lett. 209:203-205(1986).

[4]
SEQUENCE OF 318-831 FROM N.A.
MEDLINE=89184647; PubMed=2928342;
Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D., Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
"Cloning of a cDNA encoding the rat high molecular weight neurofilament peptide (NF-H): developmental and tissue expression in the rat, and mapping of its human homologue to chromosomes 1 and 22";
Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).

- FUNCTION: Neurofilaments usually contain three intermediate filament proteins: L, M, and H which are involved in the maintenance of neuronal caliber. NF-H has an important function in mature axons that is not subverted by the two smaller NF proteins. There are a number of repeats of the tripeptide K-S-P. NFH is phosphorylated on a number of the serines in this motif. It is thought that phosphorylation of NFH results in the formation of interfilament cross bridges that are important in the maintenance of axonal caliber.
- PTM: Phosphorylation seems to play a major role in the functioning of the larger neurofilament polypeptides (NF-M and NF-H), the levels of phosphorylation being altered developmentally and coincident with a change in the neurofilament function.
- SIMILARITY: Belongs to the intermediate filament family.
- CAUTION: Ref.1 sequence differs from that shown due to a frameshift in position 783.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M37227; AAA41693.1; ALT_FRAME.
EMBL; X13804; CAA32038.1; ALT_FRAME.
EMBL; M21964; AAA41695.1; --
EMBL; J04517; AAA41692.1; --
PIR; S02003; S02003.
InterPro; IPR001664; IF.
Pfam; PF00038; filament; 1.
PROSITE; PS00226; IF; 1.
InterMediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
NON_TER 1 1
DOMAIN 276 641
CONFLICT 164 164
CONFLICT 185 185
CONFLICT 193 193
CONFLICT 199 199
CONFLICT 346 346
CONFLICT 373 373
CONFLICT 482 482
CONFLICT 485 485
CONFLICT 570 571
CONFLICT 591 591
CONFLICT 727 727
CONFLICT 757 757
CONFLICT 769 769
CONFLICT 775 775
CONFLICT 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;
SEQUENCE

Query Match 7.2%; Score 241.5; DB 1; Length 831;
Best Local Similarity 22.4%; Pred. No. 0.007;
Matches 131; Conservative 87; Mismatches 199; Indels 169; Gaps 26;
QY 168 QALSNQAQKVKLEELKKKKVYVYKVEKQQAALVQLQA-ACEKREOLEHRLRLRLREL 226
DB 19 QAQAQAEAR-----DALK-----CDVTSALRRLRAQDEGHTVQSTLOSEWFRVLDRLS 68
QY 227 ESLRIQQRQGNQCFNTNVEVNAAMALLREKEERILALEA-----DMTKWEQKYLEENV 281
DB 69 EAAKVTNDAMESAQSEITEYA-----RQLQARTTLEALKSTKESLERQRSELED-- 118
QY 282 MEHPALDAAATVAARQDTTVTISHSPNTSYDTALEARLQKSEEEILMANKCLDME----- 336
DB 119 -RH-----QVDMASYQD--AIQQDNLNRLNEMAAQLREYQDLNFMALDLEIAAYR 170
QY 337 -----GRITLHAQI--IENDAMIKVLQORSRKE-----PSKTE 368
DB 171 KLLBGEBCRIGFGPIPTSLTGLPKIPSMSTHIKVKSEKIKVV-EKSEKETVIVERQTE 229
QY 369 QLSCMRPAAKSLMSISNAGSGLLSHSSTLTGSPINER--KRDDKSWKG--SLGILLGGDYR 424
DB 230 EIQ-----VTEVTEBEDKEAQGESEEBEAEEGEEA 260
QY 425 AEYVPTSPVPPTPLLSAHSKGTGRDCSTQTERGTESN-KTAAVAPISVPAVAAAAT 483
DB 261 ATTSPAEAAASP-----EKETKSPVKEEKSPABAKSPAEAKSP 300
QY 484 AAAITATATATITTTWAAAPVAV-----AAAAAATAAAT 530
DB 301 AEA--KSPAEVKSPAVAKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVK 358
QY 531 SPAAAGQIPA-----AASVASAAVAPSAARAAAVQVAPA--APAPVAPALV----- 576
DB 359 SPAEA-KSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVK 417
QY 577 -----PVPAAPAAQASAPAQTAPTS-----APAVAPTAPPTETPAVQAQVPA 620
DB 418 KSPAEVKSPVTVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVKSPAEVK 477
QY 621 SPATGPGPHRLSLPSLTCNPKDTGDPVHFHNTLTKTPIQLQGE 665
DB 478 KSPAGAKPPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSP 523

RESULT 12

ID	CPN	DRONE	STANDARD;	PRT;	865 AA.
AC	Q02910;				
DT	01-OCT-1993	(Rel. 27, Created)			
DT	01-OCT-1993	(Rel. 27, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Calphorin.				
GN	CPN OR CAP.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Spheroidea; Drosophilidae; Drosophila.				
NC	NCBI_TaxID=7227;				
RN	[1]_SEQUENCE FROM N.A.				
RC	STRAIN=Canton-S;				
RC	MEDLINE=93165729; PubMed=8094559;				
RA	Martin J.H., Benzer S., Rudnicka M., Miller C.A.;				
RT	"Calphorin: a Drosophila photoreceptor cell calcium-binding protein.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).				
RN	[2]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=Canton-S;				
RC	MEDLINE=93165730; PubMed=8434015;				
RA	Ballingner D.G., Xue N., Harshman K.D.;				
RT	"A Drosophila photoreceptor cell-specific protein, calphorin, binds				

243 VSEYNAALMELLERKEERILALEAD-----MTWEQKYLEN--VMRHFDALDAATV 293
RA Sanelin A., Schneider C., Sempie C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
324 ETEHEEPQWELLAKHETITTSREADGNTDIEKQFLDBYTKENQKVEESQADBERGEN 383
RA Urtado R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wuiming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
345 AQRDITVISHSPNTSYDTALEARIQKEBEILMANRCCLDMGRKTLHAQ----- 345
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayateu N.,
384 VAESSEIGYHEDREGND-----EKEBEDSEERNRAALREMAKLSGASRGAPVGF 437
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
346 -----IIEKDMIKVLQORSKPSKTEQLSCMRPA----- 376
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yaunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
348 NPFQMASGVGNKSEEPKQKHKEBEPEQLQELPRALPMPVPDPSSNPPFRKSNLS 497
RA Birney E., Hayashizaki Y.,
377 -----KSLMSISNAGSLLSHSTLTSTGSPIMEEKR 406
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).
CC -!- FUNCTION: Binds phospholipid vesicles containing
CC phosphatidylinositol 3-phosphate and participates in endosomal
CC trafficking (By similarity).
CC -!- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and
CC RAB22A that have been activated by GTP-binding (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
CC associated with early endosomes (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK046231; BAC32647.1; -
DR MGI; 2442192; Beal.
KW Coiled coil. 1 789 COILED COIL (POTENTIAL).
FT DOMAIN 180 550 GLN/GLU/LYS-RICH.
FT NON TER 790 790
SQ SEQUENCE 790 AA; 90539 MW; A167AB02CC2CAYBF CRC64;
Query Match 7.1%; Score 239; DB 1; Length 790;
Best Local Similarity 20.3%; Pred. No. 0.0083;
Matches 127; Conservative 96; Mismatches 187; Indels 216; Gaps 22;
QY 15 PADPFAIVSRAQM---VEILSDENRHLRQELE----- 44
DB 124 PQEVAVYVQETQKLKGSINBLTKQNLTKLQKLDYTHLEEKNEESASRKTQASL 183
QY 45 -----GCYKVARLQKVETETQVSEAYENLVKSSSKREALEKAMRNKLEGEIRMDFN 99
DB 184 HQRLDCCQLQARLTASESSLQRAQ-----ELSEKAEAAQ-----KLREELREVEST 232
QY 100 RDLRERLETANKLAKEKEYEGSEDTTKTISQIFAKNKVESOREKLEAELEATARNEDQ 159
DB 233 QHLKVEVKQLQQQREKEQGLQ-LOGEVSQLCHKLLETERQLGEARGLKQORLSSEK 291
QY 160 RRHTEIRDOALSNAQKVKVLEELKKK-----QVYVDVKVKKQOQALVOLQACEK 210
DB 292 ---LMEKEQVQADLQKLRLLEEQKEKVTNTELOHLEKSKQOHOEQOALQOSATAKL 348
QY 211 RE---QLEHLR-----TKLEHELESL---RIQROGNOCPFN 242
DB 349 REAQNDLEOVLRQIGDKQDKQTONLEALLQKESVSLLEKEREYAKIQAGEGETAVLN 408
QY 243 -VSEYNAALMEL-----LREKERRILALEADMT 270
DB 409 QLOEKHQAQOOLQTLTEKLNQSESHQABENLHDQVQKAKHLRAAQORVLSLETSVS 458
QY 271 -----KWEQKYLENVMRHFDALDAATVAAQR----- 297
DB 459 ELSQSNESKEKVSQIDQIKAKTELLSREAAKAAQADLQNHLDTAQHALQKQOELN 528
QY 298 -----DTVTISHSPNTSYDTALEARIQKEBEILMANRCCLDMGRKTLHAQIIIE--- 348

SUIT 14

A1 MOUSE

Q8BL56

10-OCT-2003 (Rel. 42, Created)

10-OCT-2003 (Rel. 42, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)

Early endosome antigen 1 (fragment).

EAL.

Mus musculus (Mouse).

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

[1]

SEQUENCE FROM N.A.

STRAIN=C57BL/6J; TISSUE=Brain;

MEDLINE=22354683; PubMed=12466851;

Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,

Nikaido I., Oeato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,

Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,

Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,

Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,

Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,

Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,

Gaasterland T., Gariboldi M., Giesi C., Godzik A., Gough J.,

Grimmond S., Gustincich S., Hachawa N., Jackson I.J., Jarvis E.D.,

Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.A.,

Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,

Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,

Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,

Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,

```

529 KVSQVQLDGLTAKFQKDEHICQLSHSLKDKHKKHLSLEQKVEDLEGHINKLEADALEVKA 598
349 -KDAIKVLYQOR-----SRKPSKTEQLSCMRPAKSLMSISNAGSLGHS 393
589 SKEQLQSLQDQQLSTDLRLNALSRELQREVVSCFK-----IDLQNKSEILENIX 643
394 STLTGSPIMEKRDCKSKMSGLGILLGDYRASYVTPSPVPSTPLLSAHSKT-----448
644 QTLT-----KKEEN-----VVLKQEFK-----LSQSKTQHKEL 674
449 GSRDCSTOTE-RGTESNKTAAVAPIS 473
675 GDRMQAAVETLTAVKAKDALLAELS 700

350T 15
180 MOUSE
1 A180 MOUSE STANDARD; PRT; 901 AA.
Q61548; O61547;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
Clathrin coat assembly protein A180 (Clathrin coat associated protein
A180) (91 kDa synaptosomal-associated protein) (Phosphoprotein P1-
20).
SNAP91.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
[1]
SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
MEDLINE=92300439; PubMed=1607933;
Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
"Characterization of a novel synapse-specific protein. II. cDNA
cloning and sequence analysis of the 21-20 protein.";
J. Neurosci. 12:2144-2155(1992).
-!- FUNCTION: Adaptins are components of the adaptor complexes which
link clathrin to receptors in coated vesicles. Clathrin-associated
protein complexes are believed to interact with the cytoplasmic
tails of membrane proteins, leading to their selection and
concentration. Binding of A180 to clathrin triskelia induces
their assembly into 60-70 nm coats.
-!- SUBCELLULAR LOCATION: Component of the coat surrounding the
cytoplasmic face of coated vesicles in the plasma membrane.
-!- ALTERNATIVE PRODUCTS:
Name=Long;
IsoId=O61548-1; Sequence=VSP 000172;
Name=Short;
IsoId=O61548-2; Sequence=VSP 000172;
-!- TISSUE SPECIFICITY: Brain. Associated with the synapses.
-!- DEVELOPMENTAL STAGE: Developmentally regulated in a pattern
coincident with active synaptogenesis and synaptic maturation.
-!- DOMAIN: Possesses a three domain structure: the N-terminal 300
residues harbor a clathrin binding site, an acidic middle domain
450 residues, interrupted by an Ala-rich segment, and the C-
terminal domain (166 residues).
-!- PTM: Phosphorylated.
-!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
EMBL; M83985; AAA37587.1; ..
EMBL; M83985; AAA37586.1; ..
PIR; A44825; A44825.
MGD; MGI:109132; Snap91.

```

```

DR InterPro: IPR001026; ENTH.
DR InterPro: IPR008943; PI_bind_N.
DR Pfam: PF01417; ENTH; 1.
DR SMART: SM00273; ENTH; 1.
DR PROSITE: PS0942; ENTH; 1.
KW Coated bits; Alternative splicing; Phosphorylation.
FT DOMAIN 14 145
FT DOMAIN 14 145
FT DOMAIN 410 413
FT DOMAIN 535 539
FT DOMAIN 547 550
FT DOMAIN 559 564
FT DOMAIN 659 664
FT DOMAIN 704 710
FT VARSPLIC 715 719
FT Missing (in isoform short).
FT /FTID=VSP_000172.
SQ SEQUENCE 901 AA; 91851 MW; 24A98FBAE8DB831 CRC64;

Query Match 7.0%; Score 234; DB 1; Length 901;
Best Local Similarity 20.8%; Pred. No. 0.015;
Matches 152; Conservative 105; Mismatches 281; Indels 192; Gaps 30;

QY 28 MVEILSDENRNLRQELGCGYKVARLQKVEIQRVSEAYENLVKSSSKRALEKAVRN- 86
DB 45 LIQATNETNVTNPQWADTLFERATNSSWV--VFKALVTTTHLMVHGNERFIQYLSRNT 102
QY 87 --KLEGEIRMHEDFNRLRLERLETANTQLEKEY-----EGSEDTKTKISQL 131
DB 103 LFNLSNPLDKSGSHCYDMSTFIRYSRYLAEKAFSYQMAFDPAVKKGAGVMTM--- 159
QY 132 FAKNKESOREKELAEALATARNEDQRRIEIRDOALSNA---QAKVVKLEELKKQ 188
DB 160 -----VPEKLLKSMPIIQG-QIDALLEFVHPNELTNGVINAAPMLLFKDLIKLFA 209
QY 189 VYVDVKEMQOALVOL-QAAEKEQLEHLRLRLERLESLRIQROGNCQPTNVSEYN 247
DB 210 CYNDGVINLEKFPKMGKQCKDALEYKFLTKMTVSEFLKVAEQVG-IDKGDIPDLT 268
QY 248 AA--ALMELLEKEERILALRADTKWEQKYLENVNRHFDALDAATVAARQDTTVISHS 305
DB 269 QAPSSIMETL---EQHLNLTLEG-----KKPGNNEGSGAPSPKSSKSPATTVT-S 313
QY 306 PNTSYDTALEARIQKEEBEILMANKRCLDMEGRIKTLHAQIIEKDAMIKVLOQSRKEPS 365
DB 314 PNST-----PACTIDTSPVDIFATASAAAP- 339
QY 366 KTEQLSCHRPKAKSLMSISNAGSLGSHSSTLTGSPIMEEKRDDKSKWGLGILLGDYRA 425
DB 340 -----VSSAKPSDLDLQPDFSGAAGNA-----APVVPSPGGATAM---GDLLGSDSLA 387
QY 426 EYVPTSPVPSTPLLSAHSKTSKGRDCSTQTERGTESNKTAAVAPISVPAVAAATAA 485
DB 388 A-LSSVPCPEAPISDPF-----APEP-SPPTTTTEPASAS 419
QY 486 ALTATATATITTMVAAPVAVAAA--AAPAAA-AAAPSPATAATAAAVSPAAAGQIPAAA 542
DB 420 ASTTTAVTAVTTEVDLFGDAFAASPGEPAPASGATAPATAPVPAALD-ACSGNDPPAP 478
QY 543 SVASAA-----AVAPSAASAAAV---QVAPAPAPVPAPALVPVPAAPAAQASAPA 590
DB 479 SSGSAAAPPELDLFAKPPETSPVPTTASTAPPVATAPSPAPTAAVATAATTAATAA 538
QY 591 QTOAPTSAPAVAPTPTP-----TPVAQAEVP-----ASPATG 625
DB 539 ATTTATTSAAATTAAPALDIFGDLFDSAPVAAAPKPDAAAPSIDLFGTDAFSSPPRG 598
QY 626 PGP-----HRLSI-----PSLTCNPKDKT-----QPVFHSNTLE-RKTPIQI 661
DB 599 ASPVPESSILADLVDAFAAPSPASTASPAKAESSGVLDLFGDAGSGASSETQAPAPQAV 658
QY 662 LQGEFDAEW 671
DB 659 SSSGASADLL 668

```

arch completed: June 16, 2004, 19:17:44
b time : 20.5037 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: June 16, 2004, 19:08:47 ; Search time 62.7139 Seconds
(without alignments)
3395.972 Million cell updates/sec

tle: US-09-332-063-2
rfect score: 3347
quence: 1 MFRAPSSASQPVADPFA.....KTPIQLGQPDADMEVYLI 675

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

al number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0
imum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SPTREMBL_25.*

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match %	Length	DB ID	Description
1	3347	100.0	675	4	Q9HDX1 homo sapien
2	2683	80.2	891	11	Q8VHC2
3	1448	43.3	956	4	Q8IY63
4	1400	41.8	882	11	Q9D4H4
5	1338	40.0	498	4	Q96CM5
6	1090.5	32.6	772	11	Q8K371
7	1087.5	32.5	463	11	Q8BS08
8	1075.5	32.1	463	11	Q8BP84
9	1069.5	32.0	586	11	Q9QU50
10	1066.5	31.9	463	11	Q7TPE4
11	1045	31.2	569	4	Q9UKB4
12	1040.5	31.1	466	4	Q96F99
13	1040.5	31.1	859	4	Q9Y2J4
14	867.5	25.9	402	4	Q8TER8
15	738.5	22.1	326	4	Q8NDN0
16	549.5	16.4	612	4	Q8TEN8

17	427.5	12.8	557	4	Q8WXD1
18	335	10.0	203	11	Q9QZP5
19	328.5	9.8	182	4	Q8H879
20	325	9.7	789	16	Q8XX15
21	310.5	9.3	3326	12	Q7T591
22	300	9.0	3288	12	Q7T5D9
23	296.5	8.9	1013	5	Q9VYT7
24	294.5	8.8	1346	5	Q9U183
25	292	8.7	507	13	Q13028
26	292	8.7	1046	11	Q8OTQ3
27	286.5	8.6	2346	5	Q01385
28	286	8.5	1069	5	Q8EBG1
29	285.5	8.5	1072	11	Q35482
30	284.5	8.5	2346	5	Q9V624
31	282	8.4	1354	11	Q9EPW8
32	280	8.4	3534	12	Q39266
33	279	8.3	801	5	Q23635
34	278	8.3	600	5	Q21036
35	275	8.2	2701	4	Q9I520
36	274.5	8.2	1299	4	Q9NSM8
37	274	8.2	426	2	Q9LAV5
38	266.5	8.0	875	2	Q9F280
39	266.5	8.0	997	5	Q9W2J2
40	266	7.9	668	5	Q8XU8
41	265	7.9	675	13	Q3W608
42	265	7.9	1020	5	Q86PC3
43	258	7.7	1174	4	Q94854
44	257.5	7.7	2037	13	Q9PV22
45	257	7.7	1480	11	Q8OTM9

ALIGNMENTS

RESULT 1

Q9HD27	PRELIMINARY;	PRT;	675 AA.
ID	Q9HD27	Q9UPT1;	
AC	Q9HD27	Q9UPT1;	
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)		
DE	Angiomotin (Hypothetical protein KIAA1071).		
GN	KIAA1071.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21157403; PubMed=11257124;		
RA	Troyanovsky B., Levchenko T., Mansson G., Matvijenko O., Holmgren L.;		
RT	"Angiomotin. An angiostatin binding protein that regulates endothelial cell migration and tube formation."		
RL	J. Cell Biol. 152:1247-1254(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=99397452; PubMed=10470851;		
RA	Kikuno R., Nagase T., Ishikawa K., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."		
RL	DNA Res. 6:197-205(1999).		
DR	ENBL; AF286598; AAG01851.1;		
DR	ENBL; A8028994; EAA8302.3;		
KW	Hypothetical protein.		
SQ	SEQUENCE 675 AA; 72540 MW; EBC28B74427AD481 CRC64;		

Query Match 100.0%; Score 3347; DB 4; Length 675;
Best Local Similarity 100.0%; Pred. No. 5.9e-138;
Matches 675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1	MPRAOPSSASQVPADPPFAIVSRAQQWVEILSDENRNLOELGCGYKVARLOKVETEI	60
1	MPRAOPSSASQVPADPPFAIVSRAQQWVEILSDENRNLOELGCGYKVARLOKVETEI	60
61	QVSVAYENLVKSSSKREALSKAMENKLEGEIRRMHDFNRDLRERLETANKOLAEKEYEG	120
61	QVSVAYENLVKSSSKREALSKAMENKLEGEIRRMHDFNRDLRERLETANKOLAEKEYEG	120
121	SDTKTTSQIPAKKESQREKELEALATARNEDORHIEIRDOALSNAQKVVKL	180
121	SDTKTTSQIPAKKESQREKELEALATARNEDORHIEIRDOALSNAQKVVKL	180
181	BEELKKQVYVDKVKMOQALVOLQAAQCEKEQLEHRLRTRLELESRLRQFGNCQP	240
181	BEELKKQVYVDKVKMOQALVOLQAAQCEKEQLEHRLRTRLELESRLRQFGNCQP	240
241	TNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVAARDTT	300
241	TNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVAARDTT	300
301	VISHSPNTSYDTALBARIQKEEBEILMANKCLDMGRIKTLHAQIIEKDMIKVLQORS	360
301	VISHSPNTSYDTALBARIQKEEBEILMANKCLDMGRIKTLHAQIIEKDMIKVLQORS	360
361	RKPSKTEQLSCWRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKRDDKSWKSGILLLG	420
361	RKPSKTEQLSCWRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKRDDKSWKSGILLLG	420
421	GVYRAEYVSTSPVPPSTPLLSAHSKTSGRDCSTQTERGTSNKTAAPVAPVAA	480
421	GVYRAEYVSTSPVPPSTPLLSAHSKTSGRDCSTQTERGTSNKTAAPVAPVAA	480
481	AATAAAITATAATITTTWVAAPVAVAAAAAFAAAAPSPATAATAAASVPAAGQIPA	540
481	AATAAAITATAATITTTWVAAPVAVAAAAAFAAAAPSPATAATAAASVPAAGQIPA	540
541	ASVASAAAFAVAAAAAFAVAPAPVAPALVPVAPAPAAQASAPACTQAPTSA	600
541	ASVASAAAFAVAAAAAFAVAPAPVAPALVPVAPAPAAQASAPACTQAPTSA	600
601	VAPTPAPTTPVAQAQEVDPASPATGPGPHRLSIPSLTCNPKDTPGVPHSNTLERTPIQ	660
601	VAPTPAPTTPVAQAQEVDPASPATGPGPHRLSIPSLTCNPKDTPGVPHSNTLERTPIQ	660
661	ILQCEPDAMWEYLI 675	
661	ILQCEPDAMWEYLI 675	
ESTLT 2		
8VHG2	PRELIMINARY; PRT; 891 AA.	
D	Q8VHG2	
C	Q8VHG2	
T	01-MAR-2002 (T-EMBLrel. 20, Created)	
T	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)	
T	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)	
E	Angiomotin.	
N	AMOT.	
S	Mus musculus (Mouse).	
C	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
C	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
X	NCBI_TaxID=10090;	
N	[1]	
P	SEQUENCE FROM N.A.	
C	STRAIN=B6CBAF2; TISSUE=Placenta;	
A	Troyanovsky B., Bratt A., Holmgren L.;	
L	"Mouse angiomotin.";	
L	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.	
R	EMBL; AF461135; AAL73436.1; -	
R	MGD; MGI:108440; Amot.	
Q	SEQUENCE 891 AA; 95111 MW; C05445424CD6088 CRC64;	
Query Match 80.2%; Score 2683; DB 11; Length 891;		

Best Local Similarity 76.8%; Pred. No. 5.5e-109;		
Matches 573; Conservative 29; Mismatches 62; Indels 82; Gaps 11;		
QY	1 MPRAOPSSASQVPADPPFAIVSRAQQWVEILSDENRNLOELGCGYKVARLOKVETEI	60
DB	157 MPRAQ-QSASYQMPADPPFAIVSRAQQWVEILSDENRNLOELGCGYKVARLOKVETEI	215
QY	61 QVSVAYENLVKSSSKREALSKAMENKLEGEIRRMHDFNRDLRERLETANKOLAEKEYEG	120
DB	216 QVSEAYENLVKSSSKREALSKAMENKLEGEIRRMHDFNRDLRERLETANKOLAEKEYEG	275
QY	121 SDTKTTSQIPAKKESQREKELEALATARNEDORHIEIRDOALSNAQ	174
DB	276 SDTKTTSQIPAKKFCQPNPAERKVVAVPLIS-----PKGTIKIRALASNAQ	329
QY	175 AKVVKLEELKKQVYVDKVKMOQALVOLQAAQCEKEQLEHRLRTRLELESRLRQOR	234
DB	330 AK-WKLEELKKQVYVDKVKMOQALVOLQAAQCEKEQLEHRLRTRLELESRLRQOR	388
QY	235 QNCQPTNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVA	294
DB	389 QGNSQPTWASEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVA	448
QY	295 AORDTTVISHSPNTSYDTALBARIQKEEBEILMANKCLDMGRIKTLHAQIIEKDMIK	354
DB	449 AORDTTVISHSPNTSYDTALBARIQKEEBEILMANKCLDMGRIKTLHAQIMEKDMIK	508
QY	355 VIQORSRKEPTEQLSCWRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKRDDKSWKGS	414
DB	509 VIQORSRKE-RETEQLSSMRPAKSLMSISNAGSGLSHSSTLTGAPIMEEKRDDKSWKGS	567
QY	415 LGILGGDYRAEYVSTSPVPPSTPLLSAHSKTSGRDCSTQTERGTSNKTAAPVAPISV	474
DB	568 LGVLLGGDYRVBPVSTSPVPPSTPLLSAHSKTSGRDCSTQTERGPESTKTAATVPIA	627
QY	475 P-APVAAAATAAAI-----TATAATITTTWVAAPVAVAAAAAFAAAAPSPATAA	524
DB	628 PMAGPVAAAAFAAANATAATNTATATNTTWTWVAAPVAVAAAAAFAAAAPSPANA	687
QY	525 ATAAA-----VSPAAAAGQIPAAAASV-----	544
DB	688 ALAAAAAATPATSVAATSVSAANSISPAAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV	747
QY	545 -----ASAAAVAPSAAAAAAFAVAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV	589
DB	748 VSPATAATAAFAAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA	805
QY	590 AQTQAPTSAPAVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV	649
DB	806 TPTQASTPAPT	865
QY	650 SNTLERTPIQLGQEPDAEMWEYLI 675	
DB	866 SSTLERKPTIQLGQEPDAEMWEYLI 891	
RESULT 3		
Q8IY63	PRELIMINARY; PRT; 956 AA.	
ID	Q8IY63	
AC	Q8IY63	
DT	01-MAR-2003 (T-EMBLrel. 23, Created)	
DT	01-MAR-2003 (T-EMBLrel. 23, Last sequence update)	
DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)	
DE	Similar to angiomotin like 1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Muscle;	
RA	Strausberg R.;	
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.	

```

EMBL: BC037539; AAH37539.1; -; 6F47AE3A13816B95 CRC64;
SEQUENCE 956 AA; 108574 MW; 43.3%; Score 1448; DB 4; Length 956;
Query Match
Best Local Similarity 62.0%; Pred. No. 2.7e-55;
Matches 292; Conservative 83; Mismatches 82; Indels 14; Gaps 6;
2 PRAOPSSAYQVPADPPAIVSRACQVVEILSDNRNLROBLECYEKVARKVQVETIQ 61
418 PQPPAASPSQQLGPDADPAIVERAQVVEILTEENRVLHQLQYNDKRLKFKELQ 477
62 RVSEAYENLVKSSKREALEKAMNKLEGEIRRMHDFNRDLRELETANKOLAEKEYEGS 121
478 RISEAYESLVKSTTKRESLDKAMNKLEGEIRRLHDFNRDLRELETANKQLSSREYEGH 537
122 EDTRKTIQSLPAKNKESQREKELEALATARSNEODRRHIEIRDOALNAQAKVYLE 181
538 ED-KAAEGHYASQNKFEKLEKLEMLAAVRTASEDHRHIEILDQALNAQAKVYLE 596
182 BELKKQVYVDKVMQOALVQLQACREKRLSHRLRLRLERLESLRIQORQNCQPT 241
597 EELREKQAYVEKVEKQALQALQALQALQALQALQALQALQALQALQALQALQALQ 656
242 NVSEYNAALMELLREKEERILALADMTKWEQKYLEBNVWHEALDAATVAQRDTTV 301
657 NNPEYAPALLELVREKEERILALADMTKWEQKYLEBESTIRHPANMAATAAERDTTI 716
302 ISHSPTSY-DTALREARIQKEEELMANKCLDMEGRITKTHAQILIEKDMIKVLQORS 360
717 INHSRNGSYGSSLEAHVQWEEVQVQANRCQDMYTIKNLHAKILIEKDMIKVLQORS 776
361 RKEPKTSQOLSMRPAKLSMISNAGSGLLSHSTLTGSPIMEKRRDKKWSKSLGILLG 420
777 RKDAGKTDSS-SLRPARSVPSIA-AATGTHSRQTSLSQALAEKKEKTKWKSIGILLG 834
421 GDYRAEYVPSPTSPVPTSP-----LSANSTKSGRDCSTQTERGTS 462
835 KEHH-EHASAPLLPPPTSSLSSTASTTAASAAKTKSGKDSSTQTKSAE 884
[1]
SEQUENCE FROM N.A.
STRAIN=CS7BL/6J; TISSUE=Testis;
MEDLINE=21085660; PubMed=11217851;
Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehi Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Pleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Moshina J., Mazzarelli J., Wembaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

```

```

Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK016526; BAB30287.1; -.
DR MGD; MGI:1922573; Anot11.
DR GO; GO:0005923; C:light function; IDA.
SQ SEQUENCE 882 AA; 98423 MW; 88AAFE83FFA56046 CRC64;
Query Match
Best Local Similarity 41.8%; Score 1400; DB 11; Length 882;
Matches 299; Conservative 80; Mismatches 103; Indels 32; Gaps 11;
1 MPRAOP-SSAYQVPADPPAIVSRACQVVEILSDNRNLROBLECYEKVAKVA 51
333 LPLPLPISLAASQPLPASPNQQLGPDADPAIVERAQVVEILTEENRVLHQLQYNDK 392
52 RLOKVETIQVSRAYENLVKSSKREALEKAMNKLEGEIRRMHDFNRDLRELETANK 111
393 KLHKEPEKLSQISSEAYESLVKSTTKRESLDKAMNKLEGEIRRLHDFNRDLRELETAN 452
112 QLAEEKEYEGSEDTRKTIQSLPAKNKESQREKELEALATARSNEODRRHIEIRDOALS 171
453 QLSSEYDGHED-KAAESHVYSQNKFEKLEKLEMLAAVRTASEDHRHIEILDQALS 511
172 NQAKVVKLEHLKKQVYVDKVMQOALVQLQACREKRLSHRLRLRLERLESLRI 231
512 NQAKVVKLEHLKKQVYVDKVMQOALVQLQACREKRLSHRLRLRLERLESLRI 571
232 QORQNCQPTVSEYNAALMELLREKEERILALADMTKWEQKYLEBNVWHEALDAATAA 291
572 QKHGTGTPVSLPECNAPALMELVREKEERILALADMTKWEQKYLEBESTIRHPAMSA 631
292 TVAAORDTTVISHPTSY-DTALREARIQKEEELMANKCLDMEGRITKTHAQILIEK 350
632 AATAERDITINHSRNGSYGSSLEAHVQWEEVQVQANRCQDMYTIKNLHAKILIEK 691
351 AMIKVLQORSRKEPKTSQOLSMRPAKLSMISNAGSGLLSHSTLTGSPIMEKRRDK- 409
692 AMIKVLQORSRKEPKTSQOLSMRPAKLSMISNAGSGLLSHSTLTGSPIMEKRRDK- 749
410 SWKSGSLGILLG-----GDYRAEYVPSPTSP---PVPSPSTELLISA-HSKTSGRDCSTQTERGT 461
750 TWKSGSIGLLGKHEHQQAAPLLPTTPASALSLEPASTTSASSTHAKTKSGKDSSTQTKST 809
462 E-----SNKTAAPVAPISVPAPVAAATAA 485
810 ELFWPSMASLPSRGRSLSTAPSNSPILKHPAAKGA 843
[1]
PRELIMINARY; PRT; 498 AA.
RESULT 5
Q96CM5 PRELIMINARY; PRT; 498 AA.
AC Q96CM5;
DT 01-DEC-2001 (TREMEL-rel. 19, Created)
DT 01-DEC-2001 (TREMEL-rel. 19, Last sequence update)
DT 01-DEC-2001 (TREMEL-rel. 19, Last annotation update)
DE Similar to RIKEN CDNA 4932416D09 gene (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC014126; AAH4126.1; -.
SQ SEQUENCE 498 AA; 56333 MW; AD557A70AB2EDD4 CRC64;
Query Match
Best Local Similarity 40.0%; Score 1338; DB 4; Length 498;
Matches 299; Conservative 80; Mismatches 103; Indels 32; Gaps 11;

```

Matches 269; Conservative 75; Mismatches 71; Indels 14; Gaps 6;

44 EGYCYKARLQKVEIQRVSAYENLVKSSKRALEKANGKLEGEIRWMDFNEDLR 103
 2 EGYVNDKHLRFEZELQRISEAYESLVKSTTKRESLDKAMRNKLEGEIRLHDFFNEDLR 61
 104 ELETANTQLAKEVEGGEDTRKTIISQLPANKEBQREKELEAEATARSTNEDQRRHI 163
 62 DLELTANROLSSREVEGHED-KAASGHYASQNKFLKEKEKLEMELEAAVRTASEDHRRHI 120
 164 ETRDQALSNAQKVKLEBEELKKQVYVDKVKQQAALVOLQAACEKREKLEHRLTRLS 223
 121 EILDQALSNAQKVKLEBEELKKQVYVDKVKQQAALVOLQAACEKREKLEHRLTRLS 180
 224 RLESIRIOROGNOGQPTNVSEYNAAALMEILREKBERILALEADMTKWEQKYLEENVMR 283
 181 RELDALRTQKNGGQANPWPYNAPALLELVREKEERILALEADMTKWEQKYLEESTIR 240
 284 HPALDAAATVAQRDTTVISHSPNTSY-DTALARIQKEEBEILMANKRCCLDMGRINKTL 342
 241 HFANMAATAAARDDTTIINHRSNGSYGESSLEAHIQEIEEVVQANRCQMEYTTIKNE 300
 343 HAQILEKDAWIKVLOORSRKESKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIM 402
 301 HAKIIEKDAWIKVLOORSRKESKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIM 358
 403 BEKRDKSKWKSIGILLGGDYRABYVPS---TPSPVPTSTPLSAHSTKSGRDC 453
 359 EERKEETKWSIGILLGKHEH-EHASAPLLPPPTSLASTTAASSAHAKTSGNDS 417
 454 STQTERGTE 462
 418 STQTDKSAE 426

RESULT 6
 8K371 PRELIMINARY; PRT; 772 AA.

AC Q8K371
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 TT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 TE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 IN Similar to angiotensin like 2.
 IS Mus musculus (Mouse).
 XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 XM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 IN NCBI_TaxID=10090;
 IP SEQUENCE FROM N.A.
 UA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 JR EMBL; BC027824; AAH27824.1; -.
 DR MGD; MGI:1929286; Amot12.
 SQ SEQUENCE 772 AA; 85278 MW; 997995F4D2A54989 CRC64;

Query Match 32.6%; Score 1090.5; DB 11; Length 772;
 Best Local Similarity 50.4%; Pred. No. 7.1e-40;
 Matches 244; Conservative 71; Mismatches 118; Indels 51; Gaps 9;

2Y 6 PSSASYQVPADPPAIV-----SRAQQWVEILSD-----ENRNLQRELECCYKVARLOK 55
 Db PSSQPPAVEGPPGAQATLGAHIAQVETVLRNARLQRDNRELORELESTSEKAGRIEK 343
 2Y 56 VETRIQVSEAVENLVKSSKRALEKANGKLEGEIRWMDFNEDLRERLETANKQALAE 115
 Db LENEIQRUSEAHESLMRTSSKRALEKANGKLEGEIRWMDFNEDLRERLETANKQALAE 403
 2Y 116 KEYEGSEDTKTIISQLPANKEBQREKELEAEATARSTNEDQRRHIETRDQALSNAQA 175
 Db KTQEAQAGSQDMWAKVLAQSYEQOQEQEKLEREMALLRGATIEDQRRHAELLEQALNAQS 463
 2Y 176 KVKVLEBEELKKQVYVDKVKQQAALVOLQAACEKREKLEHRLTRLERELRESLRIQORQ 235

Db 464 RAARAEELRRKQAVYVVERLQALQQAACREKRELETRLEQLKALRAQORQ 523
 QY 236 -----GNCQPTNVSEYNAAALMEILREKBERILALBADMTKWEQKYLEENVMRHFALDA 289
 Db 524 TOTLAGGSGHGGSHLSALRLSQLEKEBEQILALEADMTKWEQKYLEERAMQFAMDA 583
 QY 290 AATVAQRDTTVISHSPNTSYDTALARIQKEEBEILMANKRCCLDMGRINKTLHAQILEK 349
 Db 584 AATAAQRDTTLIRHSPQSPSSSP-----NSGLIPGNHRHQEKESRLKVLHAQILEK 636
 QY 350 DAMIKVLOORSRKESKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIMEKRDCK 409
 Db 637 DAVIKVLOORSRKESKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIMEKRDCK 676
 QY 410 SWKSLGILLGGDYRABYVPS---TPSPVPTSTPLSAHSTKSGRDCSTQTERGTENKT 466
 Db 677 GWQG---LVSSERQTDARPDAGDRVPAEPEPATAPLPAHTKHGSRDGTQTD-GPADNTS 731
 QY 467 AAVA 470
 Db 732 ACIA 735

RESULT 7
 Q8BS08 PRELIMINARY; PRT; 463 AA.

AC Q8BS08
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 TT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Leman coiled-coil protein.
 GN AMOTL2
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 PC STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL; AK040312; BAC30740.1; -.
 DR PIR; P70534; P70534.
 DR MGD; MGI:1929286; Amot12.
 SQ SEQUENCE 463 AA; 51758 MW; 2F7B52FBC8B1678 CRC64;

Query Match 32.5%; Score 1087.5; DB 11; Length 463;
 Best Local Similarity 52.7%; Pred. No. 5.6e-40;
 Matches 236; Conservative 68; Mismatches 103; Indels 41; Gaps 7;

QY 32 LSENENLQRELECCYKVARLOKVEIQRVSAYENLVKSSKRALEKANGKLEGEIR 91
 Db LQDNRELORELESTSEKAGRIEKLENEIQRUSEAHESLMRTSSKRALEKANGKLEGEIR 70
 QY 92 IRRHMDFNEDLRERLETANKQALAEYEGSEDTKTIISQLPANKEBQREKELEAEAT 151
 Db MRQLQDNRELORELESTSEKAGRIEKLENEIQRUSEAHESLMRTSSKRALEKANGKLEGEIR 130
 QY 152 ARSTNEQRRHIETRDQALSNAQKVKVLEBEELKKQVYVDKVKQQAALVOLQAACEKR 211
 Db LGAEIDQRRHAELLEQALNAQSRARAEELKKQVYVDKVKQQAALVOLQAACEKR 190
 QY 212 POLHRLTRLERELRESLRIQORQ-----GNCQPTNVSEYNAAALMEILREKBERILAL 265
 Db POLHRLTRLERELRESLRIQORQ-----GNCQPTNVSEYNAAALMEILREKBERILAL 250
 QY 191 EQLRLTRLERELRESLRIQORQ-----GNCQPTNVSEYNAAALMEILREKBERILAL 250
 Db EQLRLTRLERELRESLRIQORQ-----GNCQPTNVSEYNAAALMEILREKBERILAL 250
 QY 266 BADMTKWEQKYLEENVMRHFALDAAATVAQRDTTVISHSPNTSYDTALARIQKEEBEEL 325

```

> 251 EADMTKWEQKYLEERAMQFAMDAATAAQAQDRTLIRHSPQPSPPSSSF-----NEGL 303
/ 326 LMANKGLDMEGRITKLHAQIEKDKAMIKVLQORSKEPSKTEQLSCMRPAKSLMSISNA 385
> 304 LPGNHRHQMESRLKVLHAQIEKDAVIKVLQORSKDPGKATQ-GTLRPAKSVPSIFAA 362
/ 386 GSGLLSHSSTLTGSPIMEEKDDKSWKSGLSGILGDDYAEYVPS---TPSPVPPSTPLL 442
> 363 AVG-----TQWQOG-----LVSSERQTDARPAGDRVPAPPEPATAPL 399
/ 443 SHAHSTGSRDCSTOTERGTSNKTAAVA 470
> 400 PAHTKHGSRDGTQTD-GPADNTSACLA 426

RESULT 8
Q9QUS0 PRELIMINARY; PRT; 463 AA.
AC Q9QUS0;
ID Q9QUS0;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Leman coiled-coil protein.
GN AMOTL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RA Kessler R., Brunet J.-F., Van Meir B.G.;
RT "Murine homolog of the human LCCP gene. Variant 1.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF175967; RAD56362.2; -
DR MGI; MGI:1929286; Amotl2.
DR PIR; P0534; PT0534.
DR MGD; MGI:1929286; Amotl2.
SQ SEQUENCE 463 AA; 51733 MW; 0394E033ACFA17E CRC64;

Query Match 32.1%; Score 1075.5; DB 11; Length 463;
Best Local Similarity 52.2%; Pred. No. 1.8e-39;
Matches 234; Conservative 70; Mismatches 103; Indels 41; Gaps 7;

> 32 LSDNENLRQELGEGYKVARLQKVTETIORVSEAYENLVKSSSRKREALEKAMNKLEGE 91
/ 11 LORDNRLQRELESTSEKAGRIEKLNEIQLSEAHESLMRTSSKREALEKTMNKMDGE 70
/ 92 IRRMHPNRLRLTANKLAKEVEGSEDTRTISOLFPAKNKESQREKLEALAT 151
/ 71 MRRLQDFNRLRLERLESANRHLASTQEAQSQDMVAKLLAQSYEQOEKLEREMAL 130
/ 152 ARSTNEDQRHIEIRDQALSNQAQKVLLEBKVKQVVDKVMQQAALVOLQAACEKR 211
/ 131 LRGAIEDQRNAELBQALGNAQSRARAAREELRKQAYVEKVERLQQAALGOLQAACEKR 190
/ 212 BQLEHLRLRLERLESIRIQORQ-----GNCQPTNVSEYNAALMELLREKKEIRIAL 265
/ 191 EQLEHLRLRLBQELKALRAQORQTGTLAGGSGHGSABLSALSLSEQLREKQIILAL 250
/ 266 EADMTKWEQKYLEENVMRHFALDAATAAQRDRTLIRHSPQPSPPSSSF-----NEGL 303
/ 251 EADMTKWEQKYLEERAMQFAMDAATAAQAQDRTLIRHSPQPSPPSSSF-----NEGL 303
/ 326 LMANKGLDMEGRITKLHAQIEKDKAMIKVLQORSKEPSKTEQLSCMRPAKSLMSISNA 385
/ 304 LPGNHRHQMESRLKVLHAQIEKDAVIKVLQORSKDPGKATQ-GTLRPAKSVPSIFAA 362
/ 386 GSGLLSHSSTLTGSPIMEEKDDKSWKSGLSGILGDDYAEYVPS---TPSPVPPSTPLL 442
> 363 AVG-----TQWQOG-----LVSSERQTDARPAGDRVPAPPEPATAPL 399

```

```

QY 443 SHAHSTGSRDCSTOTERGTSNKTAAVA 470
DB 400 PAHTKHGSRDGTQTD-GPADNTSACLA 426

RESULT 9
Q9QUS0 PRELIMINARY; PRT; 586 AA.
AC Q9QUS0;
ID Q9QUS0;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Angiomotin-like protein 2 variant 1 (fragment).
GN AMOTL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RA Kessler R., Brunet J.-F., Van Meir B.G.;
RT "Murine homolog of the human LCCP gene. Variant 1.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF175967; RAD56362.2; -
DR MGI; MGI:1929286; Amotl2.
DR PIR; P0534; PT0534.
DR MGD; MGI:1929286; Amotl2.
DR NON_TER; -
SQ SEQUENCE 586 AA; 64946 MW; 4103BEA23B7CE613 CRC64;

Query Match 32.0%; Score 1969.5; DB 11; Length 586;
Best Local Similarity 49.8%; Pred. No. 4.3e-39;
Matches 241; Conservative 71; Mismatches 121; Indels 51; Gaps 9;

> 6 PSSASVQVPADPPFAIV-----SRAQWVEILSD-----ENRNLQELGEGYKVARLQK 55
/ 98 PSSFPQPAVEGPPSAQATLGSALHAQMTVLKRNALQDNRERLQRELESTSEKAGRIEK 157
/ 56 VETIORVSEAYENLVKSSSRKREALEKAMNKLEGEIRRMEDFNRLRLERLETANKOLAE 115
/ 158 LENEIQLSEAHESLMRTSSKREALEKTMNKMDGEMLRQDFNRLRLERLESANRHLAS 217
/ 116 KEYGSEDTRTISOLFPAKNKESQREKLEALATASTNEDQRHIEIRDQALSNQAQ 175
/ 218 KTOEAXAGSQDMVAKLLAQSYEQOEKLEREMALRGALIEDQRHAELELQALGNAQS 277
/ 176 KVKLEELKQVVDKVMQQAALVOLQAACEKRQOLEHLRLRLERLESIRIQORQ 235
/ 278 RAAREELRKQAYVERVERLQQAALGOLQAACEKGEQLEHLRLRLERLESIRIQORQ 337
/ 236 -----GNCQPTNVSEYNAALMELLREKKEIRIALADMTKWEQKYLEENVMRHFALDA 289
/ 338 TGTLAGGGSGHGSABLSALSLSEQLREKKEIQLALEADMTKWEQKYLEERAMQFAMDA 397
/ 290 AATVAQRDRTLIRHSPQPSPPSSSF-----NEGLPQNHQKESRLKVLHAQIEK 349
/ 398 AATAAQRDRTLIRHSPQPSPPSSSF-----NEGLPQNHQKESRLKVLHAQIEK 450
/ 350 DAMIKVLQORSKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKDDK 409
/ 451 DAVIKVQORSKDPGKATQ-GTLNPAKSVPSIFAAAVG-----TQ 490
/ 410 SWKSGLSGILGDDYAEYVPS---TPSPVPPSTPLLSAHSTGSRDCSTOTERGTSNKT 466
/ 491 GMDG-----LVSSERQTDARPAGDRVPAPPEPATAPLPAHTKHGSRDGTQTD-GPADNTS 545
/ 467 AAVA 470
/ 546 ACLA 549

RESULT 10
Q7TPE4 PRELIMINARY; PRT; 463 AA.
ID Q7TPE4

```



```

32  LSDNRNLROELGECYKVARLOKVEIQRVGEAVENLVKSSKEALEKAMRNKLEGE 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
11  LQDNRLQRELESSEAKAGIKLESEIQRLSEAHESLTRASKEALEKTNKNGDSE 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
92  IRMHDFNRDLRERLESTANKOLAKVEGSEDTRKTIISQLPAKVKSORKEKLELAT 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
71  MRQLQDFNRDLRERLESANRLASKTQEAQAGQDMVAKLLAQSYBQQOEKLEREMAL 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
152  ARSTNEDORRHIEIRDQALSNAAQKVVKLEELKKQVYVDKVKMQQALVOLQAACEKR 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
131  LRGAIEORRAELLEQALGNAQGRARAABEELKKQAYVEKVERIQQALQOLQAACEKR 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
212  EQLERLETRERLESRLIQORQ-----GNCQPTNVSEYNAALMELLREKEERILALE 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
191  EQLERLRLTRLEQELKALRAQQRQAGAPGGSSGGSPELSALRLSEQLREKEEQILALE 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
267  ADMTKWQKYLEENVMRHPALDAATVAARDTTVLISHSPNTSYDTALBARLIQKEEEL 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251  ADMTKWQKYLEENVMRHPALDAATVAARDTTVLISHSPNTSYDTALBARLIQKEEEL 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327  MANRCLDMEGRIKTLHAQIIEKDAMIKVLQORSRKEPSKTEQLSCNRPAKSIMSISNAG 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
304  TCGHRHQEMESLKVLAHQIIEKDAMIKVLQORSRKEPSKTEQLSCNRPAKSIMSISNAG 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
387  SGLSHSSTLTGSPIMEKRDKSKWKSGLIGLGGDYRAEYVPS-----TPSPVPSPS 438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
363  AA-----GTQWQOG-----LSSERQTDAPARLTTRAPTEPVTAT 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
439  TPLLSAHSKTSRDCSTQTERGTESNTAAVAPISVPAP-----VAAATAAATAAT 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
401  PP-AAHAKHSGRDSQVTEGPPDSTSTCL-----PEPDSLLGCSSQRAASLDVATS 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

SULT 13

```

Y2J4
Q9V2J4 PRELIMINARY; PRT; 859 AA.
01-NOV-1999 (TREMELrel. 12, Created)
01-NOV-1999 (TREMELrel. 12, Last sequence update)
01-NOV-2003 (TREMELrel. 23, Last annotation update)
Hypothetical protein KIAA0989 (Fragment).
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=99246063; PubMed=10231032;
Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
Miyajima N., Tanaka A., Kofani H., Komura N., Ohara O.,
"Prediction of the coding sequences of unidentified human genes. XIII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.",
DNA Res. 6:63-70(1999).
EMBL; AB023206; BAA76833.1; -.
Genew; HGNC:17812; AMOTL2.
InterPro; IPR00408; Reg_chromodomain.
PROSITE; PS00826; RCCL_2; 1.
Hypothetical protein.
NON_TER 1
SEQUENCE 859 AA; 94325 MW; 5AF7E211E4CF73BD CRC64;
Query Match 31.1%; Score 1040.5; DB 4; Length 859;
Best Local Similarity 46.6%; Pred. No. 1:2e-37;
Matches 237; Conservative 80; Mismatches 127; Indels 65; Gaps 10;
13 PVPDPFATVSRQQWVEI-----LSDNRNLROELGECYKVARLOKVEIQRV 63
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
375 PVSAQASSATSSGSAHLAQMEAVLRNARLQDNRLQRELESSEAKAGRIKLESRIORL 434
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14

```

Q8TER8 PRELIMINARY; PRT; 402 AA.
AC Q8TER8;
DT 01-JUN-2002 (TREMELrel. 21, Created)
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUES=pancreas;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025981; AAH25981.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 402 AA; 44234 MW; 5AD4ADF0E8A83676 CRC64;
Query Match 25.9%; Score 867.5; DB 4; Length 402;
Best Local Similarity 47.3%; Pred. No. 1:7e-30;
Matches 200; Conservative 62; Mismatches 106; Indels 55; Gaps 9;
89 EGIEIRMHDFNRDLRERLESTANKOLAKVEGSEDTRKTIISQLPAKVKSORKEKLEAE 148
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 DSEMRQLQDFNRDLRERLESANRLASKTQEAQAGQDMVAKLLAQSYBQQOEKLERE 63
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
149 LATASTNEDORRHIEIRDQALSNAAQKVVKLEELKKQVYVDKVKMQQALVOLQAAC 208
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 MALLRGATIEDQRRRAELLEQALGNAQGRARAABEELKKQAYVEKVERIQQALQOLQAAC 123
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 EKRQLEHLRLTRERLESRLIQORQ-----GNCQPTNVSEYNAALMELLREKEERIL 263
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 EKRQLEHLRLTRERLESRLIQORQAGAPGGSSGGSPELSALRLSEQLREKEEQIL 183
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
264 ALEADMTKWEQKYLEENVMRHPALDAATVAARDTTVLISHSPNTSYDTALBARLIQKEE 323
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

184 ALEADMTKEQKYLEERAMROFAMDAATAAAQDRTTLIRHSPQSPSSSF-----NE 236
324 EILMANKKCLDMEGRIKTLHAQIKEDAMIKVLOQSRKPKSKTEQLSCWRPAKSLMSIS 383
237 GLTGTGHRHQEMESLKVILHAQIKEDAMIKVLOQSRKPKSKTEQLSCWRPAKSLMSIS 295
384 NAGSGLLSHSSTLTGSPIMKRDKDKWKGSLGILLGDYRAEYVPS-----TPSPV 435
296 AAAAA-----GTQWQG-----LSSSERQTADAPARLTTRAPTEEPV 333
436 PPSPELLSAHSKTSRDCSTOTERCTESKNTAAVAPISVPAP-----VAAAATAAAITAT 490
334 VTAPP--AAHAKHGRSGSTQDGPDPSTSTCL-----PPEPDSLLGSSSQRRAASLDV 386
491 AAT 493
387 ATS 389

3SULT 15
3NDNC
3QNDNO PRELIMINARY; PRT; 326 AA.
3QNDNO;
3 01-OCT-2002 (TREMELrel. 22, Created)
3 01-OCT-2002 (TREMELrel. 22, Last sequence update)
3 01-OCT-2002 (TREMELrel. 22, Last annotation update)
3 Hypothetical protein (fragment).
3 DKFZF434O2216.
3 Homo sapiens (Human).
3 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
3 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
3 NCBI_TaxID=9606;
3 [1]
3 SEQUENCE FROM N.A.
3 TISSUE=Testis;
3 Koehrer K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
3 Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
3 EMBL; AL833833; CAD38693.1; -
3 Hypothetical protein.
3 NON_TER
3 SEQUENCE 326 AA; 35950 MW; 5E28183735E67B56 CRC64;

Query Match 22.1%; Score 738.5; DB 4; Length 326;
Best Local Similarity 59.4%; Pred. No. 5.5e-25;
Matches 152; Conservative 47; Mismatches 44; Indels 13; Gaps 5;

Y 217 RLRTLRERLESIRIQORCGNCOPTNVSEYNAALMELLREKERILLALHADMTKWEQKY 276
b 2 RLRTLRERLDALRTQQKHGNGQPNMPEYNAPALLELVREKERILLALHADMTKWEQKY 61
Y 277 LBNVVRHFAALDAATVAARQRTTVISHSPNTSY-DTALAEARIQKEEERILMANKRCLDM 335
b 62 LBNSTIRHPANNAATAAARERTTIINHSRNGSYGESSLEAHIWQEEVYQANRRQCDM 121
Y 336 EGRKITLHAQIKEDAMIKVLOQSRKPKSKTEQLSCWRPAKSLMSISNAGSGLLSHSST 395
b 122 EYTIKTLHAKIIEKAMIKVLOQSRKPKSKTEQLSCWRPAKSLMSISNAGSGLLSHSST 179
Y 396 LTGSPIMKRDKDKWKGSLGILLGDYRAEYVPSPTSPVPPSTPL-----LSAHS 446
b 180 LTFSQLAEKKEKTKWKGSLGILLGKEHH-EHASAPLPPPTLSALSSIASTTAASSAHA 238
Y 447 KTGSRDCSTOTERGTE 462
b 239 KTGSRDCSTOTDKSAE 254

```

Search completed: June 16, 2004, 19:14:52
 Job time : 65.7139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: June 16, 2004, 19:06:57 ; Search time 17.8313 Seconds
(without alignments)
2365.920 Million cell updates/sec

itle: US-09-332-063-4

irect score: 663
quence: 1 BSNKTAAPVAPISVPAPVAAA.....QASAPATQAPTAPAVAPT 143

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

imum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : A.Geneseq 29Jan04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB	ID	Description
1	663	100.0	143	3	AY54054	Angiostat
2	663	100.0	479	5	ABP43965	Abp43965 Unidentif
3	663	100.0	675	3	AY54053	AY54053 A variant
4	663	100.0	675	3	AY54052	AY54052 An angiot
5	647	97.6	993	7	AAE37918	AAE37918 Human CGD
6	226	34.1	354	6	ABU21693	ABU21693 Protein e
7	218.5	33.0	309	7	ABO23523	ABO23523 Pseudomon
8	218.5	32.9	103	5	ABP78540	ABP78540 Ala-Pro-A
9	215.5	32.5	2701	6	ABR92087	ABR92087 Human cer
10	215.5	32.5	2819	4	AAE35408	AAE35408 Human Q7C
11	209	31.5	1259	6	ABR57561	ABR57561 Human MC2
12	205.5	31.0	842	4	ABE66831	ABE66831 Drosophil
13	205.5	31.0	864	4	ABE71319	ABE71319 Drosophil
14	205.5	31.0	1743	6	ABU88255	ABU88255 Novel hum
15	205.5	31.0	1743	6	ABU90134	ABU90134 Novel hum
16	205.5	31.0	1743	6	ABU96436	ABU96436 Novel hum
17	205.5	31.0	1743	6	ABU99045	ABU99045 Novel hum
18	205.5	31.0	1743	6	ABU98260	ABU98260 Novel hum
19	205.5	31.0	1743	6	ABU91966	ABU91966 Novel hum
20	205.5	31.0	1743	6	ABU85270	ABU85270 Novel hum
21	205.5	31.0	1743	6	ABO00409	ABO00409 Novel hum
22	205.5	31.0	1743	6	ABU88960	ABU88960 Novel hum
23	205.5	31.0	1743	6	ABO06456	ABO06456 Novel hum
24	205.5	31.0	1743	6	ABU95516	ABU95516 Novel hum
25	205.5	31.0	1743	6	ABU95206	ABU95206 Novel hum

26	205.5	31.0	1743	6	ABU90754	ABU90754 Novel hum
27	205.5	31.0	1743	6	ABU93916	ABU93916 Novel hum
28	205.5	31.0	1743	6	ABU86190	ABU86190 Novel hum
29	205.5	31.0	1743	6	ABU82045	ABU82045 Novel hum
30	205.5	31.0	1743	6	ABU07906	ABU07906 Novel hum
31	205.5	31.0	1743	6	ABU94226	ABU94226 Novel hum
32	205.5	31.0	1743	6	ABO00099	ABO00099 Novel hum
33	205.5	31.0	1743	6	ABU87110	ABU87110 Novel hum
34	205.5	31.0	1743	6	ABU91351	ABU91351 Novel hum
35	205.5	31.0	1743	6	ABU90444	ABU90444 Novel hum
36	205.5	31.0	1743	6	ABU97035	ABU97035 Novel hum
37	205.5	31.0	1743	6	ABO05231	ABO05231 Novel hum
38	205	30.9	329	6	ABU39775	ABU39775 Protein e
39	204.5	30.8	1013	4	ABE71039	ABE71039 Drosophil
40	201.5	30.4	806	4	ABG04954	ABG04954 Novel hum
41	200.5	30.2	442	2	AAW26540	AAW26540 Trypanoso
42	200.5	30.2	442	2	AY23307	AY23307 Trypanoso
43	200.5	30.2	2724	4	ABG20119	ABG20119 Novel hum
44	200	30.2	477	4	ABE59024	ABE59024 Drosophil
45	197	29.7	262	2	AAW06913	AAW06913 T. cruzi

ALIGNMENTS

RESULT 1
AY54054
ID AAY54054 standard; protein; 143 AA.

XX
AC AAY54054;
XX
DT 27-MAR-2000 (first entry)
XX
DS Angiostatin-binding domain of ABP-1, designated Big-3.
XX
KW Human; angiogenesis-associated protein; plasminogen; ABP-1;
KW kringle domain; angiotensin; plasminogen receptor;
KW angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
KW inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
KW obesity; gene therapy; angiotensin-binding domain; Big-3.
XX
OS Homo sapiens.
XX
PN MO9966038-AL.
XX
PD 23-DEC-1999.
XX
PF 11-JUN-1999; 99WO-EP004109.
XX
PR 15-JUN-1998; 98SE-00002130.
PR 15-JUN-1998; 98US-0089266P.
PR 17-DEC-1998; 98SE-00004372.
PR 23-DEC-1998; 98US-0114386P.
XX
PA (PHAA) PHARMACIA & UPJOHN AB.
XX
PI Holmgren L, Troyanovsky B;
XX
PS WPI; 2000-106099/09.
PT Novel human protein useful for treating angiogenesis associated diseases
or disorders.
XX
PS Claim 6; Page 49; 58pp; English.
CC The present sequence represents the angiotensin-binding domain,
CC designated Big-3, of a human angiogenesis-associated protein. The protein
CC binds an N-terminal fragment of plasminogen. The ABP-1 protein is
CC designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or
CC kringle 5 (K5) of plasminogen. These four kringle domains comprise
CC angiotensin. The ABP-1 protein acts as a receptor for plasminogen. A
CC polymorphic variant of ABP-1 is also described, in AAY54053. ABP-1 can be
CC used to manufacture medicaments for treating angiogenesis-related

The invention relates to 446 newly isolated polynucleotide sequences. The activity of polynucleotides of the invention may be described as, vulnary, neuroprotective, immunomodulator, cytostatic and anti-inflammatory. Compositions comprising nucleic acids of the invention are useful for treating a mammalian subject, or as nutritional sources or supplements. These are useful in gene therapy, particularly for treating wounds, burns or ulcers, Alzheimer's disease, Huntington's disease.

PS Claim 5; Page 46-49; 58pp; En

PS Claim 5; Page 46-49; 58pp; En

The present sequence represents a polymorphic variant of a human angiogenesis-associated protein which is able to bind an N-terminal fragment of plasminogen. The native protein, described in AAY54052, is designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or kringle 5 (K5) of plasminogen. These four kringle domains comprise angiotensin. The ABP-1 protein acts as a receptor for plasminogen. The angiotensin-binding domain of the ABP-1 protein is described in AAY54054. ABP-1 can be used to manufacture medicaments for treating angiogenesis-related diseases or disorders, such as tumor conditions, diabetes, rheumatoid arthritis, and even some inflammatory diseases such as psoriasis, chronic inflammation of the intestine, asthma, etc. The ABP-1 protein may also be able to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene therapy techniques

Sequence 675 AA;

Query Match 100.0%; Score 663; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. No. 2.2e-38;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ESNKTAAPVAPISVPAPVAAAATAAATATATATTTTVAAPVAVAAAAAPAAAAAPSPA 60
462 ESNKTAAPVAPISVPAPVAAAATAAATATATTTTVAAPVAVAAAAAPAAAAAPSPA 521
61 TAAATAAASVPAAGQIPAAASVASAAVAPSAASAAAQVAPAPVAPALVPVPAP 120
522 TAAATAAASVPAAGQIPAAASVASAAVAPSAASAAAQVAPAPVAPALVPVPAP 581
121 AAAQASAPAQTOAPTSAPAVAPT 143
582 AAAQASAPAQTOAPTSAPAVAPT 604

RESULT 4
AAY54052
AAY54052 standard; protein; 675 AA.
AAY54052;
27-MAR-2000 (first entry)
An angiogenesis-associated protein which binds plasminogen.
Human; angiogenesis-associated protein; plasminogen; ABP-1;
kringle domain; angiotensin; plasminogen receptor;
angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
obesity; gene therapy.
Homo sapiens.
WO9966038-A1.
23-DEC-1999.
11-JUN-1999; 99WO-EP004109.
15-JUN-1998; 98SE-00002130.
15-JUN-1998; 98US-0009266P.
17-DEC-1998; 98SE-00004372.
29-DEC-1998; 98US-0114386P.
(PHAA) PHARMACIA & UPJOHN AB.
Holmgren L, Troyanovsky B;
WPI; 2000-106099/09.
N-PSDB; AA245329.
Novel human protein useful for treating angiogenesis associated diseases or disorders.

Claim 4; Page 43-46; 58pp; English.

The present sequence represents a human angiogenesis-associated protein which is able to bind an N-terminal fragment of plasminogen. The protein is designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or kringle 5 (K5) of plasminogen. These four kringle domains comprise angiotensin. The protein acts as a receptor for plasminogen. The angiotensin-binding domain of the ABP-1 protein is described in AAY54054. A polymorphic variant of ABP-1 is also described, in AAY54053. ABP-1 can be used to manufacture medicaments for treating angiogenesis-related diseases or disorders, such as tumor conditions, diabetes, rheumatoid arthritis, and even some inflammatory diseases such as psoriasis, chronic inflammation of the intestine, asthma, etc. The protein may also be able to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene therapy techniques

Sequence 675 AA;

Query Match 100.0%; Score 663; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. No. 2.2e-38;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ESNKTAAPVAPISVPAPVAAAATAAATATATTTTVAAPVAVAAAAAPAAAAAPSPA 60
462 ESNKTAAPVAPISVPAPVAAAATAAATATATTTTVAAPVAVAAAAAPAAAAAPSPA 521
61 TAAATAAASVPAAGQIPAAASVASAAVAPSAASAAAQVAPAPVAPALVPVPAP 120
522 TAAATAAASVPAAGQIPAAASVASAAVAPSAASAAAQVAPAPVAPALVPVPAP 581
121 AAAQASAPAQTOAPTSAPAVAPT 143
582 AAAQASAPAQTOAPTSAPAVAPT 604

RESULT 5
AAE37918
ID AAE37918 standard; protein; 993 AA.
XX AAE37918;
AC AAE37918;
XX 06-NOV-2003 (first entry)
XX Human CGDD-7 protein.
XX Human; cell growth, differentiation and death protein; CGDD; leukaemia;
neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
muscular disorder; myotonic dystrophy; catatonia; endocrine disorder;
diabetes; Grave's disease; cancer; immunological disorder; scleroderma;
systemic lupus erythematosus; allergy; Crohn's disease; renal disorder;
gastrointestinal disorder; Goodpasture's syndrome; infection; cirrhosis;
cardiovascular disorder; atherosclerosis; hepatic disease; transgenic;
transgenic animal; gene therapy; neuroprotective; relaxant; cytostatic;
dermatological; immunosuppressive; cerebroprotective; anticonvulsant;
antibacterial; antiparasitic; fungicide; virucide; uropathic; cardiant;
protozoacide; nootropic.
XX Homo sapiens.
XX WO2003050253-A2.
XX 19-JUN-2003.
XX 04-DEC-2002; 2002WO-US039133.
XX 07-DEC-2001; 2001US-0340747P.
XX 20-DEC-2001; 2001US-0342761P.
XX 15-JAN-2002; 2002US-0349705P.
XX 06-FEB-2002; 2002US-0354764P.
XX 12-FEB-2002; 2002US-0356216P.
(INCY-) INCYTE GENOMICS INC.

I Griffin JA, Rangkumar J, Emerling BM, Kable AB, Elliott VS;
Marquis JP, Bauchner WR, Goryad AE, Yue H, Lee EA, Becha SD, Tang YT;
I Tran UK, Swarnakar A, Lee S, Ison CH, Hafalia AJA, Tran B;
I Sprague WW, Lee ST, Khare R, Gandhi AR, Gietzen KJ, Bhatia U;
X Burrill JD, Blake JJ, Ho A, Zheng W;
X WPI; 2003-532903/50.
R N-PSDB; AAD57227.
X New CGDD polypeptides, useful for diagnosing, preventing, and treating
X disorders associated with an abnormal expression or activity of CGDD,
X e.g. neuromuscular, immunological, cardiovascular disorders, cancer
X and/or infections.
X Claim 1; Page 220-222; 29pp; English.
X The present invention relates to novel cell growth, differentiation and
X death (CGDD) proteins and polynucleotides encoding them. The sequences of
X the invention are useful in diagnosing, preventing and treating disorders
X associated with an abnormal expression or activity of CGDD such as
X neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's
X disease), muscular disorders (e.g. myotonic dystrophy, catatonias),
X endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g.
X leukemia, cervical or breast cancers), immunological disorders (e.g.
X scleroderma, systemic lupus erythematosus, allergies), gastrointestinal
X disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's
X syndrome), infections (e.g. viral, bacterial, fungal, parasitic,
X protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis)
X and hepatic diseases (e.g. cirrhosis). The polynucleotides can be used to
X create humanised animals or transgenic animals to model human diseases.
X The invention is also used in gene therapy. The present sequence is human
X CGDD-7 protein
X
X Sequence 993 AA;
Query Match 97.6%; Score 647; DB 7; Length 993;
Best Local Similarity 97.9%; Pred. No. 4.1e-37;
Matches 140; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Y 1 ESNKTAAPVAPISVPAAATAAATATATATATTTTAAATVAVAAAPVAAAPSP 60
b 780 ESNKTAAPVAPISVPAAATAAATATATATATTTTAAATVAVAAAPVAAAPSP 839
Y 61 TAAATAAASVPAAGQIPAAASVAAAPVAAAPVAAAPVAAAPVAAAPVAAAP 120
b 840 TAAATAAASVPAAGQIPAAASVAAAPVAAAPVAAAPVAAAPVAAAPVAAAP 999
Y 121 AAAQAPAPQAPTSAPAVAPT 143
b 900 AAAQAPAPQAPTSAPAVAPT 922
RESULT 5
ABU21693
ID ABU21693 standard; protein; 354 AA.
X
X ABU21693;
X
X 19-JUN-2003 (first entry)
X
X Protein encoded by prokaryotic essential gene #7220.
X
X Antisense; prokaryotic essential gene; cell proliferation; drug design.
X Burkholderia fungorum.
X WO200277183-A2.
X
X PN
X PD
X PF
X PP
X PQ
X PR

PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA25563.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 49617; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway; (8)
XX required for proliferation, or that inhibits cellular proliferation; (9)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pat_sequences
XX
XX Sequence 354 AA;
Query Match 34.1%; Score 226; DB 6; Length 354;
Best Local Similarity 46.4%; Pred. No. 2.5e-08;
Matches 64; Conservative 18; Mismatches 52; Indels 4; Gaps 3;
QY 5 TAAVAPISVPAAATAAATATATATTTTAAATVAVAAAPVAAAPVAAAPSPATAA 63
Db 77 TAAATAAASVPAAGQIPAAASVAAAPVAAAPVAAAPVAAAPVAAAPVAAAP 136
QY 64 TAAATAAASVPAAGQIPAAASVAAAPVAAAPVAAAPVAAAPVAAAPVAAAP 123
Db 137 TAAATAAASVPAAGQIPAAASVAAAPVAAAPVAAAPVAAAPVAAAPVAAAP 194
QY 124 QASAPAPQAPTSAPAVAPT 141
Db 195 GA-ATARTTRPGCAAGAA 211
RESULT 7
ABO23523
ID ABO23523 standard; protein; 309 AA.

X ABO23523;
C 04-SSP-2003 (first entry)
X Pseudomonas aeruginosa outlier protein #3.
E
W Candidate protein identification; pathogen; anti-infective;
W outlier protein; virulence protein; antigen; drug target protein;
W pathogenic organism; antimicrobial.
X
X Pseudomonas aeruginosa.
S
X US2003039963-A1.
X
X 27-FEB-2003.
X
X 30-MAR-2001; 2001US-00820843.
X
X 30-MAR-2001; 2001US-00820843.
X (BRAH/) BRAHMACHARI S K.
X (RAMA/) RAMACHANDRAN S.
X (NAND/) NANDI T.
X (BHIM/) BHIMARAO C.
X
X Brahmachari SK, Ramachandran S, Nandi T, Bhimarao C;
X WPI; 2003-492159/46.
X
X Identifying candidate proteins useful as anti-infectives involves
X matching outlier protein sequences with protein sequences in databases.
X
X Example 7; Page 37-38; 117pp; English.
X
X The present invention relates to a method for identifying candidate
X proteins in pathogens useful as anti-infectives. The invention discloses
X a computational method which involves the calculation of several sequence
X attributes and their subsequent analysis results in the identification
X of outlier proteins in different pathogens. The method is useful for the
X identification of outlier proteins (e.g. virulence proteins, antigens or
X proteins used as drug targets) in pathogenic organisms. The method of the
X invention provides reproducible results as it does not depend on the
X variable biochemical characterisation of proteins. ABO23500-ABO23617
X represent outlier proteins identified from different pathogenic organisms
X
X Sequence 309 AA;
Query Match 33.0%; Score 218.5; DB 7; Length 309;
Best Local Similarity 45.0%; Pred. No. 7.2e-08;
Matches 63; Conservative 8; Mismatches 62; Indels 7; Gaps 3;
Y 4 KTAAGVAPISVPAPVAAATATAATATTITTTVAAPVAVAAAAAPAAAAAPSPATAA 63
D 159 KTAAGVAPVAPVAAATATAATATTITTTVAAPVAVAAAAAPAAAAAPSPATAA 215
Y 64 ATAAVSPAAAGQIPAAASVAGAAVAPSAAAAAVQVAPAA-PAPVPAPALVVPVAPAA 122
D 216 AAKAAKPAK--PAKAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 272
Y 123 AQASAPACQTCAPTSAPAVAP 142
D 273 AKPAAPAASSAPAAATP 292
RESULT 8
BB78540
D ABB78540 standard; protein; 103 AA.
X ABB78540;
X
X 09-JUL-2002 (first entry)
X

DE Ala-Pro-Ala-Pro construct related protein sequence SEQ ID NO:240.
XX Plant; Gum arabic glycoprotein; GAGP; hydroxyproline-rich glycoprotein;
KW HRGP; repetitive proline-rich protein; RRRP; arabinogalactan protein;
KW AGP; plant gum.
XX
OS Acacia senegal.
OS Synthetic.
XX
PN WO200178503-A2.
XX
XX 25-OCT-2001.
PD
XX 12-APR-2001; 2001WO-US012336.
PF
XX 12-APR-2000; 2000US-00547693.
PR
XX (UYOH-) UNIV OHIO.
PA
XX Kieliszewski MJ;
PI
XX WPI; 2002-041307/05.
DR N-PSDB; ABL51829.
DR
XX Nucleic acids and proteins useful for producing hydroxy-proline rich
XX glycoproteins in plants.
PT
XX Example 25; Fig 18; 326pp; English.
PS
XX The present invention describes synthetic genes encoding plant gums and
XX other hydroxyproline (Hyp)-rich glycoproteins (HRGPs) and the nucleic
XX acids that encode them. The nucleic acids, proteins and methods from the
XX present invention may be used to produce HRGPs, repetitive proline-rich
XX proteins (RRPs) and arabinogalactan-proteins (AGPs) in plants via
XX recombinant methodologies. Also described is the expression of synthetic
XX genes designed from repetitive peptide sequences, such as glycoproteins
XX (including the peptide sequences of gum arabic glycoprotein (GAGP)).
XX ABL51730 to ABL51849 and ABB78401 to ABB78544 represent sequences used in
XX the exemplification of the present invention
SQ Sequence 103 AA;
Query Match 32.9%; Score 218; DB 5; Length 103;
Best Local Similarity 57.1%; Pred. No. 2.8e-08;
Matches 60; Conservative 3; Mismatches 36; Indels 6; Gaps 4;
OY 39 AAAPVAVAAAPAAAPSPATAATATAATATAATATAATATAATATAATATAATATAAT 98
DB 1 APAP-AP 55
OY 99 VQVAPA-APAPVPAPALVVPVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 142
DB 56 PAP 100
RESULT 9
ABR92087
ID ABR92087 standard; protein; 2701 AA.
XX
XX ABR92087;
AC
XX 10-SEP-2003 (first entry)
DT
XX Human cervical cancer cell marker encoding cDNA SEQ ID NO:83.
DE
XX Human; cervical cancer; cervical cancer marker; cancer therapy;
KW detection; gene therapy; vaccine.
KW
XX Homo sapiens.
OS
XX WO2002101075-A2.
PN
XX 19-DEC-2002.
PD

DR N-PSDB; ACC80175.
 XX New MC21 gene, useful for diagnosing or treating allergic
 PT hypersensitivity, asthma, seasonal rhinitis, urticaria, atopic dermatitis
 PT or mastocytosis.
 XX
 PS Claim 17; Page 101-106; 120pp; English.
 XX
 CC The present invention relates to novel human sequences associated with
 CC mast cell activation (ACC80172-ACC80179 and ABB57558-ABB57555). The
 CC coding sequences are useful for diagnosing or treating allergic
 CC hypersensitivity, asthma, seasonal rhinitis, urticaria, atopic dermatitis
 CC or mastocytosis
 XX
 XX Sequence 1259 AA;

Query Match 31.5%; Score 209; DB 6; Length 1259;
 Best Local Similarity 39.3%; Pred. No. 1.2e-06;
 Matches 59; Conservative 23; Mismatches 54; Indels 14; Gaps 6;
 QY 2 SNKTAIV---APISVPVAAATAAATAAATTTTMTAAAPVAVAAAAAPAAAAAPS 58
 DB 465 SSASAQVLTSAVASVLAPALASSPSSA--PTSATTSTSSPTSAP-APAPTSAPSTTPAPA 521
 QY 59 PATAATAAAVSPAAGQIPAAASVASAAVAPSAASAAAQVQVAP---AAPA--PVPAPA 113
 DB 522 PSPAAATPAPAV---PVPVLTTPSPALTPVPTPALSPAPTPAPTAAASPALTPVPTPA 578
 QY 114 LVPVPAPAAQASAPAOQAPTSAPAVAPT 143
 DB 579 LSPAPTALTPAASPALTPVPTPALSPAPT 608

RESULT 12
 ABB66631
 ID ABB66631 standard; protein; 842 AA.
 AC ABB66631;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 26685.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 DS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 FR 23-MAR-2000; 2000US-0191637P.
 XX
 FR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 DR N-PSDB; ABL10734.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 26685; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 842 AA;

Query Match 31.0%; Score 205.5; DB 4; Length 842;
 Best Local Similarity 36.9%; Pred. No. 1.5e-06;
 Matches 69; Conservative 12; Mismatches 55; Indels 51; Gaps 8;
 QY 6 RAVADISVPADPVAAATAAATAAATTAATTTTMTAAAPVAVAAAAAPVAVAAA 48
 DB 16 AVPTPSVAAAPQVWVSPAAPVAPAPAPAPAPAPAPAPAPPTLASVQVATVTPAPAPIA-AA 74
 QY 49 AAAAAA---SPATAAATAAAVSPAAGQIPAAAS-----VASAAAVAPSAAS- 96
 DB 75 VAPVASVAPPVVAATTPPAASPVSTPPVAVAQIPVAVSAPVAPVVAATPTVAVIPVAAAP 134
 QY 97 --AAVQVAPAPAP-----VPALVFPVAP-AAAQASAP--AQTOAPT 135
 DB 135 VIATPFAASAPTPAAATVPVSPVATPPVVPANTTVPVAAAPVAAVPAAPVPAVLAPA 194
 QY 136 SAPAVAP 142
 DB 195 VAPAVAP 201

RESULT 13
 ABB71319
 ID ABB71319 standard; protein; 864 AA.
 AC ABB71319;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 40749.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 DS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 FR 23-MAR-2000; 2000US-0191637P.
 XX
 FR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 DR N-PSDB; ABL15422.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 40749; 21pp + Sequence Listing; English.

XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention


```

PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090439P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090441P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98MO-US019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100844P.
PR 17-SEP-1998; 98US-0100915P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100949P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.

PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103258P.

Query Match 31.0%; Score 205.5; DB 6; Length 1743;
Best Local Similarity 57.6%; Pred. No. 2.9e-06;
Matches 57; Conservative 6; Mismatches 35; Indels 1; Gaps 1;

QY 5 TAATAVPSVPAPVAAAA-TAAATATATATTTTAAAPVAVAAAAAPAAAAAPSPATAA 63
DB 1645 TAAACATTCATATAAACTAAATATATGATTCAAAAAATAAAAAAATAAAAAA 1704

QY 64 ATAAVSPAAAGQIPAAASVASAAVAPSAASAAAVQVA 102
DB 1705 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1743

RESULT 15
ABU90134
ID ABU90134 standard; protein; 1743 AA.
AC ABU90134;
XX
DT 11-AUG-2003 (first entry)
XX
DE Novel human secreted and transmembrane PRO protein #3.
KW Human; gene therapy; tissue typing; tumour; chondrocyte proliferation;
KW chondrocyte differentiation; tumour necrosis factor-alpha release;
KW affinity purification.
XX
OS Homo sapiens.
XX
PN US2003036147-A1.
XX
PD 20-FEB-2003.
XX
PF 02-JUL-2002; 2002US-00187741.
XX
PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-00656120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069135P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078866P.

```

PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082588P.
PR 21-APR-1998; 98US-0082589P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0083322P.
PR 28-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 06-MAY-1998; 98US-0084366P.
PR 07-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 15-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 03-JUN-1998; 98US-0087759P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088555P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 11-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 02-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091496P.
PR 02-JUL-1998; 98US-0091526P.
PR 02-JUL-1998; 98US-0091528P.
PR 02-JUL-1998; 98US-0091632P.
PR 04-AUG-1998; 98US-0094006P.
PR 10-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095988P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097552P.
PR 26-AUG-1998; 98US-0097554P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098860P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98US-0101751P.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.

```

PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103258P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-00168978.

Query Match      31.0%; Score 205.5; DB 6; Length 1743;
Best Local Similarity 57.6%; Pred. No. 2.9e-06;
Matches 57; Conservative 6; Mismatches 35; Indels 1; Gaps 1;

QY      5 TAAVAPISVEAPVAAA-TAAATATATATTTTAAAPVAVAAAPAAAPSPATAA 63
Db      1645 TAAACATTCATTAATACTAAATATATGATTCATAAAAAAAAAAAAAAAAAA 1704

QY      64 ATAAVSPAAAGQIPAAASVASAAVAAPSAAAAA VOVA 102
Db      1705 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1743

```

Search completed: June 16, 2004, 19:12:40
 Job time : 20.8313 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

XM protein - protein search, using sw model

Run on: June 16, 2004, 19:08:12 ; Search time 5.94377 Seconds
(without alignments)
1242.060 Million cell updates/sec

Title: US-09-332-063-4

Perfect score: 663

Sequence: 1 ESNKTAAPISVPAPVAAA.....QASAPAQTAAPSAPAVAPT 143

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PCUTS_COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfilea1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	203.5	30.7	316	4	US-09-252-991A-32957
2	200.5	30.2	442	3	US-08-834-306-52
3	200.5	30.2	442	3	US-08-993-674A-52
4	200.5	30.2	442	4	US-09-256-976-52
5	198.5	29.9	219	2	US-08-257-309B-54
6	197	29.7	282	1	US-08-403-379A-1
7	197	29.7	282	2	US-08-929-414-1
8	197	29.7	282	2	US-08-557-309B-51
9	197	29.7	283	3	US-08-834-306-51
10	197	29.7	263	3	US-08-993-674A-51
11	197	29.7	263	4	US-09-256-976-51
12	195	29.4	180	6	5273901-7
13	195	29.4	180	6	5482703-6
14	195	29.4	805	3	US-09-103-429A-4
15	195.5	28.0	786	3	US-09-103-429A-3
16	181	27.3	399	4	US-09-252-991A-22853
17	177	26.7	2972	4	US-09-579-181-2
18	177	26.7	3118	4	US-09-579-181-1
19	175	26.4	98	2	US-07-814-220-2
20	175	26.4	98	2	US-07-812-421-2
21	168	25.3	756	4	US-09-963-137-184
22	165	24.9	304	4	US-09-252-991A-23116
23	162	24.4	206	4	US-08-529-055-54
24	162	24.4	8991	4	US-08-714-741-32
25	159.5	24.1	550	4	US-09-616-289-47
26	159	24.0	178	4	US-09-252-991A-20280
27	158.5	23.9	538	4	US-09-616-289-43

28 156.5 21.6 486 1 US-08-450-360-2 Sequence 2, Appli
29 156 23.5 750 4 US-09-165-239A-4 Sequence 4, Appli
30 153 23.1 165 4 US-09-489-039A-9067 Sequence 9067, Ap
31 152.5 23.0 618 4 US-09-252-991A-27666 Sequence 27666, A
32 150.5 22.7 195 4 US-09-252-991A-20967 Sequence 20967, A
33 149.5 22.5 183 4 US-08-529-055-50 Sequence 50, Appli
34 149 22.5 1652 4 US-09-627-650B-1 Sequence 1, Appli
35 149 22.5 1652 4 US-09-436-063C-1 Sequence 1, Appli
36 147 22.2 92 4 US-09-344-529-2 Sequence 2, Appli
37 147 22.2 193 4 US-08-529-055-49 Sequence 49, Appli
38 145.5 21.9 223 3 US-09-095-855-201 Sequence 201, App
39 145.5 21.9 223 4 US-09-205-426-201 Sequence 201, App
40 144.5 21.8 214 1 US-08-217-327-4 Sequence 4, Appli
41 144.5 21.8 229 4 US-09-252-991A-29247 Sequence 29247, A
42 143 21.6 219 4 US-09-527-345-2 Sequence 2, Appli
43 143 21.6 642 4 US-09-489-039A-12434 Sequence 12434, A
44 1079 4 US-09-489-039A-7502 Sequence 7502, Ap
45 142.5 21.5 272 4 US-09-252-991A-18313 Sequence 18313, A

ALIGNMENTS

RESULT 1

US-09-252-991A-32957
; Sequence 32957, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32957
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32957

Query Match 30.7%; Score 203.5; DB 4; Length 316;
Best Local Similarity 43.9%; Pred. No. 7.2e-09;
Matches 61; Conservative 8; Mismatches 57; Indels 13; Gaps 4;
Qy 4 KTAAVAPISVPAPVAAAATAAATATATTTTWWAAAPVAVAAAAAPAAAAAPSATAA 63
Db 174 KTAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 230
Qy 64 ATAAVSPRAAGTIPAAASVASAAAVAPASAAAQVAPAPAPVAPALVFPAPAAA 123
Db 231 AAXAAKAPAK---PRAAKPAAATAAKPAAKPA---KPAKAKPA---AKKPAKPA 280
Qy 124 QASAPAQTAAPSAPAVAP 142
Db 281 KPAAPAASSAPAPAPATP 299

RESULT 2

US-08-834-306-52
; Sequence 52, Application US/08834306
; Patent No. 6054135
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/834,306
 FILING DATE: 15-APR-1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.422C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 52:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 442 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-08-834-306-52

Query Match 30.2%; Score 200.5; DB 3; Length 442;
 Best Local Similarity 45.8%; Pred. No. 1.7e-08;
 Matches 65; Conservative 6; Mismatches 64; Indels 7; Gaps 5;

2Y 2 SNKTAAPVPSVPAVAAATATAATTTTWWAAAPVAVAAAPAAAAAPSPAT 61
 Db 298 SGKSKAKA--AAPAK-AAAPAKAAAPPAKAAAP--AKAAAPPAKAAAPPAKT 352
 2Y 62 AAATA-AAVSPAAAGQIPAAASVASAAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
 Db 353 AAPAKTAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAP--AKAAAP 411
 2Y 121 AAQAASAPAQTOAPTSAPAVAP 142
 Db 412 PAKAAAPPAKAAAPPAKAAAP 433

RESULT 3
 US-08-993-674A-52
 ; Sequence 52, Application US/08993674A
 ; Patent No. 6228372
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Smith, John M.
 ; APPLICANT: McNeill, Patricia D.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
 ; OF SEQUENCES: 81
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/993,674A
 FILING DATE: 18-DEC-1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.422C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 52:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 442 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-08-993-674A-52

Query Match 30.2%; Score 200.5; DB 3; Length 442;
 Best Local Similarity 45.8%; Pred. No. 1.7e-08;
 Matches 65; Conservative 6; Mismatches 64; Indels 7; Gaps 5;

QY 2 SNKTAAPVPSVPAVAAATATAATTTTWWAAAPVAVAAAPAAAAAPSPAT 61
 Db 298 SGKSKAKA--AAPAK-AAAPAKAAAPPAKAAAP--AKAAAPPAKAAAPPAKT 352
 QY 62 AAATA-AAVSPAAAGQIPAAASVASAAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
 Db 353 AAPAKTAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAP--AKAAAP 411
 QY 121 AAQAASAPAQTOAPTSAPAVAP 142
 Db 412 PAKAAAPPAKAAAPPAKAAAP 433

RESULT 4
 US-09-256-976-52
 ; Sequence 52, Application US/09256976
 ; Patent No. 6419933
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Smith, John M.
 ; APPLICANT: McNeill, Patricia D.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
 ; OF SEQUENCES: 210121.422C3
 ; FILE REFERENCE: 210121.422C3
 ; CURRENT APPLICATION NUMBER: US/09/256,976
 ; CURRENT FILING DATE: 1999-02-24
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 52
 ; LENGTH: 442
 ; TYPE: PRT
 ; ORGANISM: Trypanosoma cruzi
 US-09-256-976-52

Query Match 30.2%; Score 200.5; DB 4; Length 442;
 Best Local Similarity 45.8%; Pred. No. 1.7e-08;
 Matches 65; Conservative 6; Mismatches 64; Indels 7; Gaps 5;

QY 2 SNKTAAPVPSVPAVAAATATAATTTTWWAAAPVAVAAAPAAAAAPSPAT 61
 Db 298 SGKSKAKA--AAPAK-AAAPAKAAAPPAKAAAP--AKAAAPPAKAAAPPAKT 352
 QY 62 AAATA-AAVSPAAAGQIPAAASVASAAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
 Db 353 AAPAKTAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAP--AKAAAP 411
 QY 121 AAQAASAPAQTOAPTSAPAVAP 142


```
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.406C1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 263 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-929-414-1

Query Match 29.7%; Score 197; DB 2; Length 262;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAAITATAATTTTWWAAAPVAVAAAAAPAAAA 55
DB 110 EDAAAAAAKQKAAAKGAAPSGKSKAKAATAPAKAAAAAPAKAAAP-AKAAAAAPAKAA 168

QY 56 APSPTAATAAATAAVSPAAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPVAPAPA 113
DB 169 APKAAAAPAKAATAPAKAAAPAKTAAAPAKAAAPAKAAAPAKAAAPAKAATAPAKAAAPAKA 228

QY 114 LVPVPAPAAASAPAOPTAPAPAVAP 142
DB 229 ---ATAPAKA-ATAPAKAAAPAKAATAP 253

RESULT 8
US-08-557-309B-51
; Sequence 51, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,309B
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-929-414-1
```

```
US-08-557-309B-51

Query Match 29.7%; Score 197; DB 2; Length 263;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAAITATAATTTTWWAAAPVAVAAAAAPAAAA 55
DB 111 EDAAAAAAKQKAAAKGAAPSGKSKAKAATAPAKAAAAAPAKAAAP-AKAAAAAPAKAA 169

QY 56 APSPTAATAAATAAVSPAAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPVAPAPA 113
DB 170 APKAAAAPAKAATAPAKAAAPAKTAAAPAKAAAPAKAAAPAKAATAPAKAAAPAKA 229

QY 114 LVPVPAPAAASAPAOPTAPAPAVAP 142
DB 230 ---ATAPAKA-ATAPAKAAAPAKAATAP 254

RESULT 9
US-08-834-306-51
; Sequence 51, Application US/08834306
; Patent No. 6054135
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,306
; FILING DATE: 15-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-834-306-51

Query Match 29.7%; Score 197; DB 3; Length 263;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAAITATAATTTTWWAAAPVAVAAAAAPAAAA 55
DB 111 EDAAAAAAKQKAAAKGAAPSGKSKAKAATAPAKAAAAAPAKAAAP-AKAAAAAPAKAA 169

QY 56 APSPTAATAAATAAVSPAAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPVAPAPA 113
DB 170 APKAAAAPAKAATAPAKAAAPAKTAAAPAKAAAPAKAAAPAKAATAPAKAAAPAKA 229
```

QY 114 LVPVPAPAAQAASAPAQTOAPTSPAPAVAP 142
| | | | | : | | | | : | | | |
DB 230 ---ATAPAKA-ATAPAKAAAPAKAATAP 254

RESULT 10
US-08-993-674A-51
; Sequence 51, Application US/08993674A
; Patent No. 6228372
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,674A
; FILING DATE: 18-DEC-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-993-674A-51

Query Match 29.7%; Score 197; DB 3; Length 263;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAATATAATTTTWWAAAPVAVAAAAAPAAAA 55
| | | | | : | | | | : | | | |
DB 111 EDAAAAAAKQKAAAKKAAAPSGKSKAAATAPAKAAAPAKAATAAP-AGAAAAAPAKAAA 169
| | | | | : | | | | : | | | |

QY 56 APSPTAATAATAAVSPAAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPAPVPAPA 113
| | | | | : | | | | : | | | |
DB 170 APAKAAAPAKAATAAPAKAAAPAKTAAAPAKAAAPAKAATAAPAKAATAAPAKAATAAPAKA 229
| | | | | : | | | | : | | | |

QY 114 LVPVPAPAAQAASAPAQTOAPTSPAPAVAP 142
| | | | | : | | | | : | | | |
DB 230 ---ATAPAKA-ATAPAKAAAPAKAATAP 254

RESULT 12
5273901-7
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOROZOITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 03-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 7;
; LENGTH: 180
5273901-7

Query Match 29.4%; Score 195; DB 6; Length 180;
Best Local Similarity 43.0%; Pred. No. 1.8e-08;
Matches 65; Conservative 8; Mismatches 66; Indels 12; Gaps 4;

QY 1 ESNKTAAVAPISVPAPVAAAATAAA-ITATAATTTTWWAAAPV---AVAAAAAPAAAA 55
| | | | | : | | | | : | | | |
DB 27 EKEEERAAAPAAATAAPAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86
| | | | | : | | | | : | | | |

QY 56 APSPTAATAATAAVSPAA-----AGQIPAAASVASAAAVAPSA--AAAAVQVAPAPAP 108
| | | | | : | | | | : | | | |
DB 87 APTAATAAAGAGAAKGAKEARTETEGAGAAABAEKAKTQAATAATTAAAAAASAAAG 146
| | | | | : | | | | : | | | |

QY 109 VPAPALVVPAPAAQAASAPAQTOAPTSPAPA 139

; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 51
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
; FEATURE:
; OTHER INFORMATION: Where any xaa is an independently selected amino
; OTHER INFORMATION: acid
US-09-256-976-51

Query Match 29.7%; Score 197; DB 4; Length 263;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAATATAATTTTWWAAAPVAVAAAAAPAAAA 55
| | | | | : | | | | : | | | |
DB 111 EDAAAAAAKQKAAAKKAAAPSGKSKAAATAPAKAAAPAKAATAAP-AGAAAAAPAKAAA 169
| | | | | : | | | | : | | | |

QY 56 APSPTAATAATAAVSPAAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPAPVPAPA 113
| | | | | : | | | | : | | | |
DB 170 APAKAAAPAKAATAAPAKAAAPAKTAAAPAKAAAPAKAATAAPAKAATAAPAKAATAAPAKA 229
| | | | | : | | | | : | | | |

QY 114 LVPVPAPAAQAASAPAQTOAPTSPAPAVAP 142
| | | | | : | | | | : | | | |
DB 230 ---ATAPAKA-ATAPAKAAAPAKAATAP 254

RESULT 12
5273901-7
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOROZOITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 03-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 7;
; LENGTH: 180
5273901-7

Query Match 29.4%; Score 195; DB 6; Length 180;
Best Local Similarity 43.0%; Pred. No. 1.8e-08;
Matches 65; Conservative 8; Mismatches 66; Indels 12; Gaps 4;

QY 1 ESNKTAAVAPISVPAPVAAAATAAA-ITATAATTTTWWAAAPV---AVAAAAAPAAAA 55
| | | | | : | | | | : | | | |
DB 27 EKEEERAAAPAAATAAPAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86
| | | | | : | | | | : | | | |

QY 56 APSPTAATAATAAVSPAA-----AGQIPAAASVASAAAVAPSA--AAAAVQVAPAPAP 108
| | | | | : | | | | : | | | |
DB 87 APTAATAAAGAGAAKGAKEARTETEGAGAAABAEKAKTQAATAATTAAAAAASAAAG 146
| | | | | : | | | | : | | | |

QY 109 VPAPALVVPAPAAQAASAPAQTOAPTSPAPA 139

db 147 KPGGHAASAAKQAQKAANAATAATA 177

RESULT 13
Patent No. 5482709-6
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
NUMBER OF SEQUENCES: 10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/148,432
FILING DATE: 08-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 581,693
FILING DATE: 12-SEP-1990
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1989
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 6:
LENGTH: 130
5482709-6

Query Match 29.4%; Score 195; DB 6; Length 180;
Best Local Similarity 43.0%; Pred. No. 1.8e-08;
Matches 65; Conservative 8; Mismatches 66; Indels 12; Gaps 4;

2y 1 ESNKTAAVAPISVPAPVAATATAA-ITATAITTTTVAAPV-----AVAAAAAPAAAA 55
db 27 EXEERAAAPAATAAAPAAATRAATAATAATPAAAAPAAAPAAATCGAA 86

2y 56 APSPTATAATAAAVSPAA-----AGOTPAASVASAAVAAPSAA-AAAVOVVAPAPAP 108
db 87 APAATATAAGAGAAGAKABARTETEGAGAGAAAEKAKTQATATTAAARAASAAG 146

2y 109 VPAPALVPVPAPAAQAAPATQAPTSPA 139
db 147 KPGGHAASAAKQAQKAANAATAATA 177

RESULT 14
US-09-103-429A-4
Sequence 4, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: CDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tioga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids

NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 805 amino acids

TYPE: amino acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-4

Query Match 29.4%; Score 195; DB 3; Length 805;
Best Local Similarity 37.7%; Pred. No. 8.4e-08;
Matches 66; Conservative 10; Mismatches 61; Indels 38; Gaps 8;

QY 5 TAAVAPISVPAPVAATATAAATAATI-----TTTTVAAAAPVAVAAA-----A 49
DB 529 TAIPTPAPTAAPTAAPTADESPTTVTVPPTAAPTAAPTTAVEITITTSAPTAAPTA 588

QY 50 ADAAA--AAP-----SPATAAATAAAVSPAAAGQIPAAAASVAGSAAVAPSA 95
DB 589 ATTAAPTAAPTTAVPEITTTVSPTTAAPTAAAPNTTVTPPTAA-PTTAADAPNTTV 647

QY 96 AAADVOP--AAPAPVPAPALVPVPAPAAAO---ASAPAQOAP---TSAPAVAP 142
DB 648 TAPPTAAPTAAPTAAPTNTTVTPPTAAPTAAAPTVAHAPNTTAAAPTPTSAPTTP 702

RESULT 15
US-09-103-429A-3
Sequence 3, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: CDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tioga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
S-09-103-429A-3

Query Match 28.0%; Score 185.5; DB 3; Length 786;
Best Local Similarity 37.6%; Pred. No. 4.4e-07;
Matches 59; Conservative 8; Mismatches 69; Indels 21; Gaps 6;
y 5 TAAVAPISVPAPVAAATAAAITATAATI-----TTMVAAAPVAVAAA--AAPAA 53
b 529 TAIPTTAPTAAPTAAPTTAAPESPTTVTPPTAAPTAAFTTAVEEIPITVTSAPTAAPTA 588
y 54 AAAPSPATAAATAAASVPAAGQIPAAASVASAAAVAPSAAAAAVQVAP--AAPAPVEA 111
b 589 APTAAPTAAPTTAUPEIPTTVTSPTAA--PTTAAPAPNTTVTPPTAAPTAAAPAPNTT 646
y 112 PALVPVPAPAAQ---ASAPAQTOAP---TSAPAVAP 142
b 647 VIVPPTAAPTAAAPTVAHAPNTTAAPVTTTSAPATTP 683

search completed: June 16, 2004, 19:13:24
db time : 6.94377 secs

Result No.	Score	Query [†]		DB	ID	Description
		Match	Length			
1	249.5	37.6	4640	14	US-10-184-644-75	Sequence 75, Appl
2	249.5	37.6	4640	14	US-10-184-634-75	Sequence 75, Appl
3	249	37.6	2846	14	US-10-184-644-169	Sequence 169, Appl
4	249	37.6	2846	14	US-10-184-634-169	Sequence 169, Appl
5	249	37.6	2846	14	US-10-053-685-37	Sequence 37, Appl
6	247	37.3	1904	12	US-10-142-626-99	Sequence 99, Appl
7	247	37.3	1904	14	US-10-123-155-99	Sequence 99, Appl
8	247	37.3	1904	14	US-10-146-731-99	Sequence 99, Appl
9	247	37.3	1904	14	US-10-140-472-99	Sequence 99, Appl
10	247	37.3	1904	14	US-10-141-761-99	Sequence 99, Appl
11	247	37.3	1904	14	US-10-142-885-99	Sequence 99, Appl
12	247	37.3	1904	14	US-10-158-790-99	Sequence 99, Appl
13	247	37.3	1904	15	US-10-137-871-99	Sequence 99, Appl
14	247	37.3	1904	15	US-10-140-523-99	Sequence 99, Appl
15	247	37.3	1904	15	US-10-141-756-99	Sequence 99, Appl

[illegible]


```

? CURRENT APPLICATION NUMBER: US/10/184,644
?
? CURRENT FILING DATE: 2002-06-28
? Prior Application removed - See File Wrapper or Palm
? NUMBER OF SEQ ID NOS: 612
? SEQ ID NO 169
? LENGTH: 2846
? TYPE: DNA
? ORGANISM: Homo sapien
US-10-184-644-169

```

```

US-10-184-644-169

Query Match      37.6%; Score 249; DB 14; Length 2846;
Best Local Similarity 50.4%; Pred. No. 1.9e-08;
Matches 68; Conservative 8; Mismatches 59; Indels 0; Gaps 0;

QY      5 TAAVAPISVPARAAATAAATTTATTAATTTTTWAAAAPVAVAAAAAAPAAAAAPSPATAAA 64
DB      2711 TACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 2770

QY      65 TAAASVSPAAGAQIPAAASVASAAVAAPSAAAAAQVAPAPAPVPVPAALVFVPAPAQAQ 124
DB      2771 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2830

QY      125 ASAPAOQTQAPTSA 139
DB      2831 AAAAAAAAAAAAAAAAAAAAA 2845

```

```

1  APPLICANT: Gurney,Austin L.
2  APPLICANT: Pan,James
3  APPLICANT: Smith,Victoria
4  APPLICANT: Watanabe,Colin K.
5  APPLICANT: Wood,William I.
6  APPLICANT: Zhang,Zemin
7  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
8  TITLE OF INVENTION: ACIDS ENCODING THE SAME
9  FILE REFERENCE: P3430R1C217
10 CURRENT APPLICATION NUMBER: US/10/184,634
11 CURRENT FILING DATE: 2002-06-28
12 Prior Application removed - See file Wrapper or Palm
13 NUMBER OF SEQ ID NOS: 612
14 SEQ ID NO 169
15 LENGTH: 2846
16 TYPE: DNA
17 ORGANISM: Homo Sapien
18 US-10-184-634-1169

```

	Query Match	37.6%	Score 249;	DB 14;	Length 2846;
	Best Local Similarity	50.4%;	Pred. NO. 1.9e-08;		
Matches	68;	Conservative	8;	Mismatches 59;	Indels 0; Gaps 0;
QY	5	TAAVAPISVPAPAAAAATAAATATTATTWVAAPVAVAAAAAAPAAAAPSPTAAA	64		
Db	2711	TACAAAAANAAA	2770		
QY	65	TAAASVPAAAGCQPAAASVASAAAAVAPSAAAAANAAYOVVAPAPAAPVFVPALUVEVPAPAAQAQ	124		
Db	2771	AAAAAAAANAAA	2830		
QY	125	ASAPAQCTQAPTSA PA	139		
Db	2831	AAAAAAAAAAAAAAAAAAAA	2845		


```

; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-871-99

Query Match      37.3%; Score 247; DB 15; Length 1904;
Best Local Similarity 53.5%; Pred. No. 1.8e-08;
Matches 68; Conservative 7; Mismatches 52; Indels 0; Gaps 0;

QY 15 APVAAATATAATATATTTTWWAAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAA 74
DB 1770 AACAAATATAAGTGAATCCCAAAAAAATAAAAAAATAAAAAAATAAAAAA 1829

QY 75 GQTPAAASVAAASAAVAPAAAAAQQVAPAPAPVPAPALVPVPAPAAQAASAPACTOAP 134
DB 1830 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 1889

QY 135 TSAPAVA 141
DB 1890 AAAAAA 1896

RESULT 14
US-10-140-923-99
; Sequence 99, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C200
; CURRENT APPLICATION NUMBER: US/10140923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 99
; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-923-99

Query Match      37.3%; Score 247; DB 15; Length 1904;
Best Local Similarity 53.5%; Pred. No. 1.8e-08;
Matches 68; Conservative 7; Mismatches 52; Indels 0; Gaps 0;

QY 15 APVAAATATAATATATTTTWWAAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAA 74
DB 1770 AACAAATATAAGTGAATCCCAAAAAAATAAAAAAATAAAAAAATAAAAAA 1829

QY 75 GQTPAAASVAAASAAVAPAAAAAQQVAPAPAPVPAPALVPVPAPAAQAASAPACTOAP 134
DB 1830 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 1889

QY 135 TSAPAVA 141
DB 1890 AAAAAA 1896
```

```

RESULT 15
US-10-141-756-99
; Sequence 99, Application US/10141756
; Publication No. US20030207359A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C200
; CURRENT APPLICATION NUMBER: US/10141756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 99
; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-756-99

Query Match      37.3%; Score 247; DB 15; Length 1904;
Best Local Similarity 53.5%; Pred. No. 1.8e-08;
Matches 68; Conservative 7; Mismatches 52; Indels 0; Gaps 0;

QY 15 APVAAATATAATATATTTTWWAAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAA 74
DB 1770 AACAAATATAAGTGAATCCCAAAAAAATAAAAAAATAAAAAAATAAAAAA 1829

QY 75 GQTPAAASVAAASAAVAPAAAAAQQVAPAPAPVPAPALVPVPAPAAQAASAPACTOAP 134
DB 1830 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 1889

QY 135 TSAPAVA 141
DB 1890 AAAAAA 1896

Search completed: June 16, 2004, 19:17:10
Job time : 14.8105 secs
```


Query Match 33.1%; Score 219.5; DB 2; Length 1299;
 Best Local Similarity 34.7%; Pred. No. 3e-05;
 Matches 59; Conservative 25; Mismatches 47; Indels 39; Gaps 5;

QY 5 TAAVAPISVPAPVAAAATA-----AAITATAATITTTWAAAPVAVAAAAAPAAAAP 57
 DB 328 TSAPVASTLAPVLASTAPVPSLAPVSASVSASVPASTGAAAITSSAPASAPAP 387
 QY 58 SP-----ATAATAAVSPAAAGQIPAAASVAVASAAVAPSA 95
 DB 388 TPIIASVSTPASTVILASASIPILASALASTSAPTPAASSPAAPVI--TAPTIPASAP 445
 QY 96 AAIVQVAA---APAPVPAPALVPVAPAAQASAPAQOTQPTSPAPAP 142
 DB 446 TASVPLAFASAPAPAPTPVSAFNEP-----PAPAQTOQTHKPAQNP 490

RESULT 3
 S29309
 C:Species: Pseudomonas aeruginosa
 C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 07-May-1999
 C:Accession: S29309
 R:Timin, A.; Steinbuechel, A.
 A:Title: Cloning and molecular analysis of the poly(3-hydroxyvalkanoic acid) gene locus
 A:Reference number: S29309; PMID:1396693
 A:Accession: S29309
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-217 <TIM>
 A:Cross-references: EXBL:X66592

Query Match 33.0%; Score 218.5; DB 2; Length 217;
 Best Local Similarity 45.0%; Pred. No. 8.3e-06;
 Matches 63; Conservative 8; Mismatches 62; Indels 7; Gaps 3;

QY 4 KTAAVAPISVPAPVAAAATAATAITATAITTTWAAAPVAVAAAAAPAAAAPSPATAA 63
 DB 67 KTAAPAKPAKPAKAAKAPAA---KPAKKTAAKTAAPKPAKPAKTAAPKPAKPAK 123
 QY 64 ATAAVSPAAAGQIPAAASVAVASAAVAPSAAAAVQVAPAA--PAPVPAPALVPVAPAA 122
 DB 124 AAKAAKAPAAK---PAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 180
 QY 123 AQASAPAQOTQPTSPAPAPAP 142
 DB 181 AKPAAPAASSAPAPAPATP 200

RESULT 4
 G83013
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Jun-2001
 C:Accession: G83013
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bhardwaj, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, N.; Lory, S.; Olson, M.V.
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
 A:Reference number: A82950; PMID:20437337; PMID:10984043
 A:Accession: G83013
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-309 <STO>
 A:Cross-references: GB:AE004919; GB:AE004091; NID:g9951346; PIDN:AG08445.1; GSPDB:GN001
 A:Experimental source: strain PAO1
 A:Genetics:
 A:Gene: phaf; PA5060

Query Match 33.0%; Score 218.5; DB 2; Length 309;

Best Local Similarity 45.0%; Pred. No. 1.1e-05;
 Matches 63; Conservative 8; Mismatches 62; Indels 7; Gaps 3;

QY 4 KTAAVAPISVPAPVAAAATAATAITATAITTTWAAAPVAVAAAAAPAAAAPSPATAA 63
 DB 159 KTAAPAKPAKPAKAAKAPAA---KPAKKTAAKTAAPKPAKPAKTAAPKPAKPAK 215
 QY 64 ATAAVSPAAAGQIPAAASVAVASAAVAPSAAAAVQVAPAA--PAPVPAPALVPVAPAA 122
 DB 216 AAKAAKAPAAK---PAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 272
 QY 123 AQASAPAQOTQPTSPAPAPAP 142
 DB 273 AKPAAPAASSAPAPAPATP 292

RESULT 5
 A47282
 C:Species: Drosophila melanogaster
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: A47282
 R:Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
 A:Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
 A:Reference number: A47282; PMID:93165729; PMID:8094559
 A:Accession: A47282
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-865 <MAR>
 A:Cross-references: GB:L02111; NID:gl57031; PIDN:AA28405.1; PID:gl57032
 A:Experimental source: photoreceptor cells
 A:Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBI:124956)
 C:Genetics:
 A:Gene: flyBase:Cpn
 A:Cross-references: FlyBase:FBgn0010218
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
 C:Keywords: calcium binding

Query Match 31.7%; Score 210; DB 2; Length 865;
 Best Local Similarity 37.8%; Pred. No. 7e-05;
 Matches 70; Conservative 14; Mismatches 53; Indels 48; Gaps 8;

QY 6 AAVALISVPAPVAAAATAATAITATAITTTWAAAPVAVAAAAAPVAVAA-- 46
 DB 16 APWTPSAVAAPVQVVSAPAAVAPAPAPAPAPAPAPAPAPPTLASVQVATVTIPAPAPAAASV 75
 QY 47 ---AAAAAPAAAAPSPATAA-----ATAAVSPAAAG-----QIPAAAASVA 84
 DB 76 TPVASVAPVAVAAAPTPPAASPVSTPVAQIPVAVAPVAPVATPTPVQIPVAAAPVI 135
 QY 85 SAAAVAPSAAAAVQ---VAP--AAPVPAPALVPVAP--AAQASAP--AQOQAPTS 137
 DB 136 ATPPVAASAPTPAAVTPVISPVIASPPVPPVANTTPVAAAPVAVAPVAVPVLAPAVA 195
 QY 138 PAVAP 142
 DB 196 PAVAP 200

RESULT 6
 T42567
 C:Species: equine herpesvirus 4
 A:Variety: strain NS80567
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T42567
 R:Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
 A:Title: The DNA sequence of equine herpesvirus-4.
 A:Reference number: Z22173; PMID:98264497; PMID:9603335
 A:Accession: T42567
 A:Status: preliminary; translated from GB/EMBL/DDBJ

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

XX protein - protein search, using sw model

run on: June 16, 2004, 19:12:42 ; Search time 3.49633 seconds

(without alignments)

2129.669 Million cell updates/sec

title: US-09-332-063-4

perfect score: 663

sequence: 1 ESNKTAAPVAPVAAA.....QASAPAQTAQTSAPAVAPT 143

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	210	31.7	865	1 CPN_DROME	Q02910 drosophila
2	207	31.2	439	1 XP2_XENLA	P17437 xenopus lae
3	192	29.0	316	1 CDNC_HUMAN	P49918 homo sapien
4	185.5	28.0	662	1 MUC1_XENLA	Q05049 xenopus lae
5	183.5	27.7	299	1 RNP2_DROME	P50887 drosophila
6	183.5	27.7	790	1 ALNP_NOTOC	P24856 notothenia
7	181	27.3	352	1 ALGP_PSEAE	P15276 pseudomonas
8	176.5	26.6	797	1 VGLX_HSVB	P28968 equine herp
9	175	26.4	546	1 NUP1_DROME	Q9vdv3 drosophila
10	168	25.3	1300	1 SAL3_HUMAN	Q9bxa9 homo sapien
11	167.5	25.3	901	1 A180_MOUSE	Q61548 mus musculu
12	166.5	25.1	915	1 A180_RAT	Q05140 rattus norv
13	166	25.0	163	1 CU38_LOCM1	P04375 locusta mig
14	166	25.0	386	1 HXAD_MOUSE	Q62424 mus musculu
15	165.5	25.0	1046	1 IF2_STRAW	Q82K53 streptomyce
16	165.5	25.0	1083	1 T2D3_HUMAN	Q00268 homo sapien
17	164	24.7	388	1 HKAD_HUMAN	P13171 homo sapien
18	164	24.7	676	1 ICPO_HSVBJ	P29128 bovine herp
19	164	24.7	907	1 A180_HUMAN	Q60641 homo sapien
20	159	24.0	477	1 HXZ_HUMAN	P56270 homo sapien
21	159.5	23.9	416	1 ODO2_ALCEU	P52973 alcaligenes
22	158	23.8	518	1 TPW4_DROME	P49455 drosophila
23	157	23.7	1033	1 IF2_STRO	Q8CJG8 streptomyce
24	156.5	23.6	699	1 VGLG_HSV2H	P13290 herpes simp
25	154	23.2	721	1 YK82_MYCTU	Q10690 mycobacteri
26	153	23.1	553	1 ODP2_ALCEU	P29098 alcaligenes
27	153	23.1	676	1 ICPO_HSVBK	P29836 bovine herp
28	152.5	23.0	1004	1 IF2_CORGL	Q8np40 corynebacte
29	152	22.9	331	1 MAZ_MESAU	P56670 mesocricetu
30	152	22.9	477	1 MAZ_MOUSE	P56671 mus musculu
31	151	22.8	361	1 IF35_MOUSE	Q9dch4 mus musculu
32	149.5	22.5	1446	1 IE18_PRVIF	P33479 pseudorabie
33	148.5	22.4	3421	1 TEGU_HSVB	P28955 equine herp

ALIGNMENTS

RESULT 1

CPN_DROME	ID	CPN_DROME	STANDARD	PRT	865 AA
34	148	22.3	3828	1	TRX_DROVI
35	147.5	22.2	672	1	HMW3_MYCPN
36	147.5	22.2	3149	1	TEGU_EBV
37	146.5	22.1	324	1	OSA_DROSI
38	146.5	22.1	466	1	HOBI_SCHPO
39	146	22.0	1783	1	RAA3_CHLRE
40	145	21.9	157	1	CU63_LOCM1
41	145	21.9	959	1	PTK1_RALSO
42	144.5	21.8	1367	1	AMVH_YEAST
43	144	21.7	1115	1	NCAL_MOUSE
44	143.5	21.6	1461	1	IE18_PRVIF
45	143.5	21.6	2716	1	OSA_DROME

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RT Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RL "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RT Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT calcium and contains a leucine zipper";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -!- FUNCTION: Might function as a calcium-sequestering "sponge" to
CC regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
CC of Ca(2+) per mole of protein.
CC -!- SUBUNIT: Homodimer (Probable).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; hypodense compartment.
CC -!- TISSUE SPECIFICITY: Soma and axons of photoreceptor cells of
CC compound eyes and ocelli.
CC -!- DEVELOPMENTAL STAGE: Expressed early in photoreceptor cell
CC development.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L02111; AAA28405.1; -;
CC EMBL; L05080; AAA28420.1; -;
CC PIR; A47282; A47282.
CC PIR; A47283; A47283.
CC Flybase; FBgn0010218; Cpn.
CC GO; GO:0005509; F:calcium ion binding; IDA.
CC Calcium-binding.
CC CONFLICT 36 36 A -> AVAPAWVA (IN REF. 2).
CC CONFLICT 43 43 I -> T (IN REF. 2).
CC

Q24742 drosophila
Q50360 mycoplasma
P01186 epstein-bar
Q26682 drosophila
O74352 schizosacch
Q9feca chlamydomon
P45584 locusta mig
Q8xrh0 talstonia s
P08640 saccharomyc
P13595 mus musculu
P11675 pseudorabie
Q8in94 drosophila

CC Name=7;
CC IsoId=Q05049-7; Sequence=VSP_004647;
CC -!- TISSUE SPECIFICITY: Skin.
CC -!- PTM: Extensively O-glycosylated.
CC -!- SIMILARITY: Contains 6 P-type (trefoil) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
XR EMBL; L03115; AA74725.1; .
XR PIR; A45155; A45155.
XR HSSP; PG1359; 2ESP.
XR InterPro; IPR000519; P_trefoil.
XR Pfam; PF00088; trefoil; 6.
XR PRINTS; PR00680; PTFREFOIL.
XR SMART; SM00018; PD; 6.
XR PROSITE; PS00025; P_TREFOIL; 6.
CW Repeat; Glycoprotein; Alternative splicing.
QT NON_TER 1 1
QT DOMAIN 81 144
TT 8 X 8 AA APPROXIMATE TANDEM REPEATS,
TT ALA/THR-RICH.
TT 1-1. 88
TT REPEAT 81 88
TT 1-2. 96
TT REPEAT 89 96
TT 1-3. 104
TT REPEAT 97 104
TT 1-4. 112
TT REPEAT 105 112
TT 1-5. 120
TT REPEAT 113 120
TT 1-6. 128
TT REPEAT 121 128
TT 1-7. 136
TT REPEAT 129 136
TT 1-8. 144
TT REPEAT 137 144
TT DOMAIN 161 202
TT 8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
TT 1-1. 201
TT REPEAT 218 201
TT 1-2. 224
TT REPEAT 225 224
TT 2-3. 240
TT REPEAT 240 249
TT 2-4. 259
TT REPEAT 250 259
TT 2-5. 275
TT REPEAT 260 275
TT 2-6. 287
TT REPEAT 276 287
TT 2-7. 294
TT REPEAT 288 294
TT 2-8. 301
TT REPEAT 295 301
TT 3-1. 347
TT DOMAIN 306 347
TT 3-2. 353
TT DOMAIN 353 394
TT 3-3. 402
TT DOMAIN 402 522
TT 3-4. 411
TT REPEAT 402 411
TT 3-5. 419
TT REPEAT 412 419
TT 3-6. 431
TT REPEAT 420 431
TT 3-7. 443
TT REPEAT 432 443
TT 3-8. 453
TT REPEAT 444 453
TT 3-9. 460
TT REPEAT 454 460
TT 3-10. 472
TT REPEAT 461 472
TT 3-11. 479
TT REPEAT 473 479
TT 3-12. 491
TT REPEAT 480 491
TT 3-13. 498
TT REPEAT 482 498
TT 3-14. 515
TT REPEAT 499 515
TT 3-15. 522
TT REPEAT 516 522
TT 4-1. 566
TT DOMAIN 525 566
TT 4-2. 613
TT DOMAIN 572 613
TT 4-3. 661
TT DOMAIN 620 661
TT 4-4. 162
TT DISULFID 162 188
TT 4-5. 172
TT DISULFID 172 187
TT 4-6. 199
TT DISULFID 182 199
TT 4-7. 307
TT DISULFID 307 333
TT 4-8. 317
TT DISULFID 317 332
TT 4-9. 344
TT DISULFID 327 344
TT 4-10. 354
TT DISULFID 354 380
TT 4-11. 364
TT DISULFID 364 379
TT 4-12. 374
TT DISULFID 374 391
TT 4-13. 526
TT DISULFID 526 552

FT DISULFID 536 551 BY SIMILARITY.
FT DISULFID 546 563 BY SIMILARITY.
FT DISULFID 573 599 BY SIMILARITY.
FT DISULFID 583 598 BY SIMILARITY.
FT DISULFID 593 610 BY SIMILARITY.
FT DISULFID 621 647 BY SIMILARITY.
FT DISULFID 631 646 BY SIMILARITY.
FT DISULFID 641 658 BY SIMILARITY.
FT VARSPLIC 240 259 Missing (in isoform 5 and isoform 6).
FT VARSPLIC 250 259 /Ftid=vsp_004646.
FT VARSPLIC 250 259 Missing (in isoform 4 and isoform 7).
FT VARSPLIC 276 294 /Ftid=vsp_004647.
FT VARSPLIC 276 294 Missing (in isoform 4 and isoform 6).
FT VARSPLIC 278 278 /Ftid=vsp_004648.
FT VARSPLIC 278 278 Missing (in isoform 5).
FT VARSPLIC 306 350 /Ftid=vsp_004649.
FT VARSPLIC 420 498 Missing (in isoform 2 and isoform 5).
FT VARSPLIC 420 498 /Ftid=vsp_004650.
FT VARSPLIC 420 498 Missing (in isoform 3).
FT VARIANT 276 276 K -> E.
FT VARIANT 354 354 C -> R.
FT VARIANT 415 415 T -> A.
SQ SEQUENCE 662 AA; 67774 MW; F085277F1ED2FD40 CRC64;
Query Match 28.0%; Score 185.5; DB 1; Length 662;
Best Local Similarity 40.4%; Pred. No. 0.0012;
Matches 57; Conservative 10; Mismatches 57; Indels 17; Gaps 4;
QY 19 AATAAAITATATITTTTVAAPVAVAAAA-----APAAAAAPSPATAATAAAYS 70
DB 1 APTTAAAVATGKDTTAAAGSAAAEKTAAGVSAPPTAAVAATGEDATTAATAAET 60
QY 71 PAAAGQIP--AAASVASAAAVAPSAASAAAVQVAPAAP-----VPAPALVVPAPAA 122
DB 61 TAAAGAPTITPATTAAGKAPTAAATAPTTA--AAGAPTATGAKAPATAAAAPVPTAA 119
QY 123 AQASAPQOAPTAPSAVAAPT 143
DB 120 SKAPTAAATHSTAAAAAPT 140
RESULT 5
RL22_DROME STANDARD; PRT; 299 AA.
AC P50887; Q9V3X9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DB 60S ribosomal protein L22.
GN RPL22 OR BG-BACR19J1.4 OR CG7434.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Glover C.V.C., Bidwai A.P., Zhao W.F.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borikova D.,
RA Minano B., Kafatos P.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikoti F.,
RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,

JA Glover D.M.;
 IT "From sequence to chromosome: the tip of the X chromosome of D.
 MT melanogaster";
 L Science 287:2220-2222(2000).
 [3]
 RP SEQUENCE FROM N.A.
 IC STRAIN=Berkley;
 K MEDLINE=20196006; PubMed=10731132;
 A Adams M.D., Celisner S.E., Holt R.W., Hoskins R.A., Galle R.F.,
 A Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 A George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 A Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 A Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 A Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 A Abril J.P., Agbayani A., An H.-U., Andrews-Pfannkuch C., Baldwin D.,
 A Balles R.M., Basu A., Bakendale J.P., Bayraktaroglu L., Beasley E.M.,
 A Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 A Borkeva D., Borchan M.R., Bouck J., Brockstein P., Brothier P.,
 A Burtis K.C., Susam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 A Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 A de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 A Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 A Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 A Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 A Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 A Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 A Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 A Jaiali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 A Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 A Laoko P., Lai Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 A Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 A Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 A Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 A Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 A Palazzolo M., Pittman G.S., Pan S., Pollard J., Puzi V., Reese M.G.,
 A Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 A Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 A Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 A Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 A Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissbach J.,
 A Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 A Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 A Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 A Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 T "The genome sequence of Drosophila melanogaster";
 L Science 287:2185-2195(2000).
 C -1- SIMILARITY: Belongs to the L22E family of ribosomal proteins.
 C
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 C
 R EMBL; U42587; AB017433.1; -
 R EMBL; ALJ32792; CAB60023.1; -
 R EMBL; AE003418; AAP45546.1; -
 R FlyBase; FBgn0015288; RpL22.
 R InterPro; IPR002671; Ribosomal_L22e.
 R Pfam; PF01776; Ribosomal_L22e; 1.
 R ProDom; PD007306; Ribosomal_L22e; 1.
 W Ribosomal protein.
 W DOMAIN 24 31 POLY-ALA.
 T DOMAIN 46 50 POLY-ALA.
 T DOMAIN 65 70 POLY-ALA.
 T DOMAIN 93 98 POLY-ALA.
 T DOMAIN 103 112 POLY-ALA.
 T DOMAIN 136 152 POLY-ALA.
 T DOMAIN 185 188 POLY-LYS.
 T DOMAIN 292 299 ASP/GLU-RICH (HIGHLY ACIDIC).
 Q SEQUENCE 299 AA; 46A99005610E4EB0 CRC64;

Query Match 27.7% Score 183.5; DB 1; Length 299;
 Best Local Similarity 40.0% Pred. No. 0.00084;
 Matches 64; Conservative 9; Mismatches 54; Indels 33; Gaps 7;
 QY 4 KTAAVADI-SVPAPVAAAA-----TAAAITATAATTITTTTAAAPVAVAAAAAPA 52
 DB 13 KTAAPKPAEKKAAPAAAAAAGKVKPKAABAPAAAAAANKVKKASAAKDVKAAAAAKP 72
 QY 53 AAAPSPATTAATA-----AAVSP-----AAAGQIPAAASVASAAAVAPSA----- 94
 DB 73 AAAPKAAKPAASAKDAGKAPAAAAAPKDDAKAAAPAPAKAAPAKKAAATPAAPAKK 132
 QY 95 AAAAVQVAPAPAPVP-----APALV-PVPAPAAQASAPAQ 130
 DB 133 AAPAKAAAPAAAAAPAAAAAPAAVAKPAPKPKAKAAPPSK 172
 RESULT 6
 ANP_NOTCO
 ID -ANP NOTCO STANDARD; PRT; 790 AA.
 AC P24856;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Antifreeze glycoprotein precursor (AFGP polypeptide)
 DE [Contains: AFGP7 (AFGP 7); AFGP8 (AFGP 8)] (Fragment).
 GN AFGP8.
 OS Notothenia coriiceps neglecta (black rockcod) (Yellowbelly rockcod).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
 OC Nototheniidae; Nototheniidae; Notothenia.
 OX NCBI_TaxID=8209;
 RN NCBI
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RE MEDLINE=91067687; PubMed=2251271;
 RA Hsiao K.-C., Cheng C.-H.C., Fernandes I.B., Detrich H.W. III,
 RA Devries A.L.;
 RA "An antifreeze glycoprotein gene from the antarctic cod Notothenia
 RA coriiceps neglecta encodes a polypeptide of high peptide copy
 RA number";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9265-9269(1990).
 RN [2]
 RP REVISIONS TO N-TERMINUS AND 457.
 RA Cheng C.-H.C.;
 RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Antifreeze proteins lower the blood freezing point.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Synthesized by the liver and secreted into
 CC the blood from which they become distributed to almost the entire
 CC extracellular space.
 CC -1- DOMAIN: Contains 44 copies of AFGP8 and two copies of AFGP7.
 CC -1- PTM: Glycosylated; contains disaccharide Galactose-N-
 CC acetylglactosamine attached to threonines in AFGP8 and AFGP7.
 CC -1- DATABASE: NAME=Protein Spotlight;
 CC NOTE=Issue 5 of December 2000;
 CC WWW="http://www.expsy.org/spotlight/articles/sptl005.html".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M55000; AAA49392.2; -
 DR PIR; A38420; A38420.
 KW Antifreeze protein; Glycoprotein; Polypeptide; Repeat;
 KW Multigene family.
 FT NON_TER 1 1

```

>T PROPEP <1 5
>T CHAIN 6 790
>T PEPTIDE 6 19 ANTIPEEZE GLYCOPEPTIDE POLYPROTEIN.
>T PEPTIDE 23 36 AFGP8.
>T PEPTIDE 40 53 AFGP8.
>T PEPTIDE 57 70 AFGP8.
>T PEPTIDE 74 87 AFGP8.
>T PEPTIDE 91 104 AFGP8.
>T PEPTIDE 108 121 AFGP8.
>T PEPTIDE 125 138 AFGP8.
>T PEPTIDE 142 155 AFGP8.
>T PEPTIDE 159 172 AFGP8.
>T PEPTIDE 176 189 AFGP8.
>T PEPTIDE 193 206 AFGP8.
>T PEPTIDE 210 223 AFGP8.
>T PEPTIDE 227 240 AFGP8.
>T PEPTIDE 244 257 AFGP8.
>T PEPTIDE 261 274 AFGP8.
>T PEPTIDE 278 291 AFGP8.
>T PEPTIDE 295 308 AFGP8.
>T PEPTIDE 312 325 AFGP8.
>T PEPTIDE 329 342 AFGP8.
>T PEPTIDE 346 359 AFGP8.
>T PEPTIDE 363 376 AFGP8.
>T PEPTIDE 380 393 AFGP8.
>T PEPTIDE 397 410 AFGP8.
>T PEPTIDE 414 427 AFGP8.
>T PEPTIDE 431 444 AFGP8.
>T PEPTIDE 448 461 AFGP8.
>T PEPTIDE 465 478 AFGP8.
>T PEPTIDE 482 495 AFGP8.
>T PEPTIDE 499 512 AFGP8.
>T PEPTIDE 516 529 AFGP8.
>T PEPTIDE 533 546 AFGP8.
>T PEPTIDE 550 563 AFGP8.
>T PEPTIDE 567 580 AFGP8.
>T PEPTIDE 584 597 AFGP8.
>T PEPTIDE 601 614 AFGP8.
>T PEPTIDE 618 631 AFGP8.
>T PEPTIDE 635 648 AFGP8.
>T PEPTIDE 652 665 AFGP8.
>T PEPTIDE 669 682 AFGP8.
>T PEPTIDE 686 699 AFGP8.
>T PEPTIDE 703 716 AFGP8.
>T PEPTIDE 720 736 AFGP8.
>T PEPTIDE 740 756 AFGP8.
>T PEPTIDE 760 773 AFGP8.
>T PEPTIDE 777 790 AFGP8-LIKE.
>T SEQUENCE 790 AA; 71265 MW; 4C7CCADC48P902 CRC64;

Query Match 27.7%; Score 183.5; DB 1; Length 790;
Best Local Similarity 40.0%; Pred. No. 0.0017;
Matches 62; Conservative 10; Mismatches 66; Indels 17; Gaps 4;

QY 5 TAAVAPISVPVAAATATAA-----ITATATATITTTWVAAPVA--VAAAAAPAAAAAPS 58
DB 184 TPATALNFATATATPATATPATATPATATPATATPATATPATATPATATPATATPATATPATAT 243
QY 59 PATATAATAAASPA-----AAGQIPAAASVASAAAVAPASAAAQVQVAPAPAPV 109
DB 244 AATAATPATATPATATPATATPATATPATATPATATPATATPATATPATATPATATPATATPATAT 303
QY 110 PAPALVFPVAPAA--AQASAAQTAQTSAPAPAP 142
DB 304 PATAALNFATATATPATATPATATPATATPATATPATATPATATPATATPATATPATATPATAT 338

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional regulatory protein algp (Alginate regulatory protein
DE algr3).
GN ALGP OR ALGR3 OR PA5253.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8822;
RX MEDLINE=90108714; PubMed=2514124;
RA Kato J., Chu L., Kitano K., Devault J.D., Kimbara K.,
RA Chakrabarty A.M., Misra T.K.;
RT "Nucleotide sequence of a regulatory region controlling alginate
RT synthesis in Pseudomonas aeruginosa: characterization of the algr2
RT gene.";
RL Gene 84:31-38(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8882;
RX MEDLINE=90222135; PubMed=2109318;
RA Kato J., Misra T.K., Chakrabarty A.M.;
RT "Algr3, a protein resembling eukaryotic histone H1, regulates
RT alginate synthesis in Pseudomonas aeruginosa.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2887-2891(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=8830;
RX MEDLINE=91008921; PubMed=1698761;
RA Deretic V., Konyeceni W.M.;
RT "A prokaryotic regulatory factor with a histone H1-like
RT carboxy-terminal domain: clonal variation of repeats within algp, a
RT gene involved in regulation of mucoidy in Pseudomonas aeruginosa.";
RL J. Bacteriol. 172:5544-5554(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO / PAO2003;
RX MEDLINE=90236911; PubMed=2110144;
RA Konyeceni W.M., Deretic V.;
RT "DNA sequence and expression analysis of algp and alqO, components of
RT the multigene system transcriptionally regulating mucoidy in
RT Pseudomonas aeruginosa: algp contains multiple direct repeats.";
RL J. Bacteriol. 172:2511-2520(1990).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman P.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: THE PROMOTER FOR A CRITICAL ALGINATE BIOSYNTHETIC
CC GENE, ALGP, ENCODING GDP-MANNOSE DEHYDROGENASE, IS ACTIVATED ONLY
CC UNDER CONDITIONS REMINISCENT OF THE CYSTIC FIBROSIS LUNG (I.E.,
CC UNDER HIGH OSMOLARITY), AND AT LEAST TWO REGULATORY GENES, ALGP
CC AND ALQO, HAVE BEEN IMPLICATED IN THIS ACTIVATION PROCESS.
CC -!- DOMAIN: THE CARBOXY-TERMINAL BINDS TO DNA. IT IS UNKNOWN WHETHER
CC BINDING IS SPECIFIC OR NON-SPECIFIC.
CC -!- SIMILARITY: TO EUKARYOTIC HISTONES H1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

RESULT 7
ALGP_PSEAE
ID ALGP_PSEAE STANDARD; PRT; 352 AA.
AC P15276; Q8HTU1;
DF 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

EMBL; M30145; AAA25724.1; --
EMBL; M35259; AAA25705.1; --
EMBL; M57551; AAA25703.1; --
EMBL; M32077; AAA72068.1; --
EMBL; AB004937; AAG08638.1; --
PIR; A35630; A35630.
PIR; G82990; G82990.
InterPro; IPRO06970; PT.
Pfam; PF0486; Pf; 3.
W Arginate biosynthesis; Transcription regulation; Activator;
DNA-binding; Repeat; Complete proteome.
T CONFLICT 28 28 G -> D (IN REF. 4).
T CONFLICT 157 158 KP -> NA (IN REF. 1 AND 2).
T CONFLICT 158 158 P -> R (IN REF. 3).
T CONFLICT 173 176 MISSING (IN REF. 1, 2 AND 3).
T CONFLICT 176 176 A -> T (IN REF. 4).
T CONFLICT 181 181 A -> G (IN REF. 3 AND 4).
T CONFLICT 188 188 T -> A (IN REF. 1, 2 AND 3).
T CONFLICT 223 224 KP -> NA (IN REF. 1 AND 2).
T CONFLICT 246 246 T -> A (IN REF. 1, 2 AND 3).
T CONFLICT 266 267 PA -> HV (IN REF. 1, 2 AND 3).
T CONFLICT 273 280 MISSING (IN REF. 1, 2 AND 3).
T CONFLICT 290 291 KP -> NA (IN REF. 1 AND 2).
T CONFLICT 311 311 T -> A (IN REF. 1, 2 AND 3).
T CONFLICT 320 321 KP -> NA (IN REF. 1 AND 2).
T CONFLICT 320 321 KP -> NA (IN REF. 1 AND 2).
Q SEQUENCE 352 AA; B70C32AE67C8012E CRC64;
Query Match 27.3%; Score 181; DB 1; Length 352;
Best Local Similarity 38.3%; Pred. No. 0.0013;
Matches 59; Conservative 6; Mismatches 71; Indels 18; Gaps 4;
Y 4 KTAAPVAPISVP-----APVAAATAAATATATATTTTAAVAAAPVAAVAAAPAA 54
b 177 KTAAPVAPISVP-----APVAAATAAATATATATTTTAAVAAAPVAAVAAAPAA 236
Y 55 AAPSP-----ATAATAAASVAPAAAGVAPAAASVAPAAASVAPAAASVAPAA 109
b 237 PAAPVAPVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAA 292
Y 110 PAPALVPVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAA 143
b 293 AKPAAPVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAAAGVAPAA 326
RESULT 8
ID VGLX HSVB STANDARD; PRT; 797 AA.
AC P28958;
XT 01-DEC-1992 (Rel. 24, Created)
XT 01-DEC-1992 (Rel. 24, Last sequence update)
XT 01-DEC-1992 (Rel. 24, Last annotation update)
XE Glycoprotein X precursor.
EN 71.
XS Equine herpesvirus type 1 (strain Ab4p) (SHV-1).
XC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
XC Alphaherpesvirinae; Varicellovirus.
XX NCBI_TaxID=31520;
XN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1";
RL Virology 189:304-316(1992).
IC This SWISS-PROT entry is copyright. It is produced through a collaboration
IC between the Swiss Institute of Bioinformatics and the EMBL outstation -
IC the European Bioinformatics Institute. There are no restrictions on its
IC use by non-profit institutions as long as its content is in no way
IC modified and this statement is not removed. Usage by and for commercial
IC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
IC or send an email to license@isb-sib.ch.)

EMBL; M86664; AAB02506.1; --
PIR; H36802; VGBEX1.
InterPro; IPRO07110; Ig-like.
XW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 797 GLYCOPROTEIN X.
FT DOMAIN 23 465 SER/THR-RICH.
FT TRANSMEM 766 790 POTENTIAL.
FT CARBOHYD 590 590 N-LINKED (GLCNAC...) (POTENTIAL).
Q SEQUENCE 797 AA; 80342 MW; 50C9ED9211F5E5E2 CRC64;
Query Match 26.6%; Score 176.5; DB 1; Length 797;
Best Local Similarity 30.9%; Pred. No. 0.0041;
Matches 51; Conservative 19; Mismatches 72; Indels 23; Gaps 4;
QY 2 SNKTAAPVAPISVPAPVA-----AAATAAATAAATIT--TTWAAAPVAAVAAAPAA 53
Db 163 TSTTTTATTTTPTTASTTTTDTTAAATTAATTAATTAATTAATTAATTAATTAATTA 222
QY 54 AAAPSPATAATAAASVAPAA--AGVAPAAASVAPAAASVAPAAASVAPAAASVAPAA 110
Db 223 ATTSATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 282
QY 111 APALVP-----VPAPAAASVAPAAASVAPAAASVAPAAASVAPAAASVAPAA 143
Db 283 ATSATPTSTSAATTTTPTTSAATSAATSAATSAATSAATSAATSAATSAATSAAT 327
RESULT 9
ID NUP1 DROME STANDARD; PRT; 546 AA.
AC Q9VDV3; Q8SYM7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable nucleoporin Nup58.
GN NUP58 OR CG7360.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter R.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jatali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,


```

DR HSP; P07248; 1A8E.
DR Genew; HGNC:10527; SALL3.
DR MIM; 605079; -.
DR InterPro; IPR007887; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2_10.
DR SMART; SM00355; Znf_C2H2_10.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW Metal-binding; Repeat; Alternative splicing; Polymorphism.
FT ZN_FING 420 442 C2H2-TYPE.
FT ZN_FING 448 470 C2H2-TYPE.
FT ZN_FING 679 701 C2H2-TYPE.
FT ZN_FING 707 729 C2H2-TYPE.
FT ZN_FING 739 761 C2H2-TYPE.
FT ZN_FING 977 999 C2H2-TYPE.
FT ZN_FING 1005 1027 C2H2-TYPE.
FT ZN_FING 1113 1135 C2H2-TYPE.
FT ZN_FING 1141 1163 C2H2-TYPE.
FT DOMAIN 198 201 POLY-ALA.
FT DOMAIN 213 216 POLY-GLN.
FT DOMAIN 397 902 POLY-SER.
FT VARSPLIC 1 133 Missing (in isoform 2 and isoform 4).
FT VARSPLIC 973 1044 /FTID=VSP_006832.
FT VARSPLIC 973 1044 Missing (in isoform 1 and isoform 2).
FT VARSPLIC 973 1044 /FTID=VSP_006833.
FT VARIANT 593 593 V -> L.
FT CONFLICT 235 235 R -> C (IN REF. 2).
FT CONFLICT 235 235 R -> C (IN REF. 2).
SQ SEQUENCE 1300 AA; 135371 MW; 82CF3BDCBD59150 CRC64;

Query Match 25.3%; Score 168; DB 1; Length 1300;
Best Local Similarity 28.8%; Pred. No. 0.016;
Matches 59; Conservative 19; Mismatches 51; Indels 76; Gaps 7;

QY 10 PISVPAPVAATAAATATATTTTVAAPVA-----VAAAAP----- 51
DB 148 PPAAPAPTPAYGAPSTNTVLEALLSTKVAQFSQGRAGGAGGVAAPVPLILE 207
QY 52 -----AAAAAPSPATAATAAAGVPA 73
DB 208 QLMALQQQIHLQLEIRSQVALMQRPSPSPSPAPSPAPSPQLGLAALPLS 267
QY 74 AGQIPAAASVASAAVAPSAASAAA-----AVQVAPAPAPVPAAPV 117
DB 268 AG-AFAAA-IAGSGPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 324
QY 118 PAPAQAASAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 142
DB 325 PAP-APQAASAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 348

RESULT 11
A180_MOUSE
ID A180_MOUSE STANDARD; PRT; 901 AA.
AC Q61548; Q61547; PRT; 901 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Clathrin coat assembly protein AP180 (Clathrin coat associated protein
DE AP180) (91 kDa synaptosomal-associated protein) (Phosphoprotein P1-
DE 20)
DE SNAP91.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RX MEDLINE=92300439; PubMed=1607933;
RA Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
RT "Characterization of a novel synapse-specific protein. II. cDNA
RT cloning and sequence analysis of the F1-20 protein.";
```

```

RL J. Neurosci. 12:2144-2155(1992).
CC -!- FUNCTION: Adaptins are components of the adaptor complexes which
CC link clathrin to receptors in coated vesicles. Clathrin-associated
CC protein complexes are believed to interact with the cytoplasmic
CC tails of membrane proteins, leading to their selection and
CC concentration. Binding of AP180 to clathrin triskelia induces
CC their assembly into 60-70 nm coats.
CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the
CC cytoplasmic face of coated vesicles in the plasma membrane.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q61548-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q61548-2; Sequence=VSP_000172;
CC -!- TISSUE SPECIFICITY: Brain. Associated with the synapses.
CC -!- DEVELOPMENTAL STAGE: Developmentally regulated in a pattern
CC coincident with active synaptogenesis and synaptic maturation.
CC -!- DOMAIN: Possesses a three domain structure: the N-terminal 300
CC residues harbor a clathrin binding site, an acidic middle domain
CC 450 residues, interrupted by an Ala-rich segment, and the C-
CC terminal domain (166 residues).
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M83985; AAA37587.1; -.
CC EMBL; M83985; AAA37586.1; -.
CC FIR; A44825; A44825.
CC MGD; MGI:109132; Snap91.
CC InterPro; IPR01026; ENTH.
CC InterPro; IPR008943; PI_bind_N.
CC Pfam; PF01417; ENTH; 1.
CC SMART; SM00273; ENTH; 1.
CC PROSITE; PS50342; ENTH; 1.
CC Coated pits; Alternative splicing; Phosphorylation.
KW DOMAIN 14 145 ENTH.
FT DOMAIN 410 413 POLY-THR.
FT DOMAIN 535 539 POLY-ALA.
FT DOMAIN 547 550 POLY-ALA.
FT DOMAIN 659 664 POLY-SER.
FT DOMAIN 704 710 POLY-SER.
FT VARSPLIC 715 719 Missing (in isoform Short).
FT VARSPLIC 715 719 /FTID=VSP_000172.
SQ SEQUENCE 901 AA; 91851 MW; 24A98F8ACE8DB8B1 CRC64;

Query Match 25.3%; Score 167.5; DB 1; Length 901;
Best Local Similarity 37.0%; Pred. No. 0.013;
Matches 60; Conservative 9; Mismatches 64; Indels 29; Gaps 6;

QY 9 APISVP-APVAAATAAATAATATTTTVAAPVAV-----AAAAAPAAAAAPS----- 58
DB 396 APISDPFAPPEPPTTTTTPASASATTTAVTTTEVDLFGDAFAASPGAPAAAPGAT 455
QY 59 -PATAATAAASVSPAAGQIPAAASVASAAAV-----APSAASAAAVQVAPAPAPVPA 111
DB 456 APATAPVPAALD-ACSGNDPFPSPGSAEAPALDLFAMKPPETSAFVVTPTTASTAPPV 514
QY 112 PALVVPAP-----AAAQASAPAQTAQTSAPAVAP 142
DB 515 PATAPSPAPTAATAATAATTTAAAAATTTTAAAAATTTTAAAAATTAAP 556

RESULT 12
A180_RAT
ID A180_RAT STANDARD; PRT; 915 AA.
```

AC Q05140;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Clathrin coat assembly protein AP180 (Clathrin coat associated protein)
 DE AP180 (91 kDa synaptonemal-associated protein).
 GN SNAP91.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RC TISSUE=Brain;
 RX MEDLINE=93178442; PubMed=8440257;
 RA Morris S.A., Schroeder S., Plessmann U., Weber K., Ungewickell E.;
 RT "Clathrin assembly protein AP180: primary structure, domain
 organization and identification of a clathrin binding site.";
 RL EMBO J. 12:667-675 (1993).
 CC -!- FUNCTION: Adaptins are components of the adaptor complexes which
 link clathrin to receptors in coated vesicles. Clathrin-associated
 protein complexes are believed to interact with the cytoplasmic
 tails of membrane proteins, leading to their selection and
 concentration. Binding of AP180 to clathrin triskelia induces
 their assembly into 60-70 nm coats.
 CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the
 cytoplasmic face of coated vesicles in the plasma membrane.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=Q05140-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=Q05140-2; Sequence=VSP_000173;
 CC -!- DOMAIN: Possesses a three domain structure: the N-terminal 300
 residues harbor a clathrin binding site, an acidic middle domain
 450 residues, interrupted by an Ala-rich segment, and the C-
 terminal domain (166 residues).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; X68977; CAA48748.1; -;
 DR EMBL; X68978; CAA48749.1; -;
 DR FIR; S36326; S36326.
 DR FIR; S36327; S36327.
 DR InterPro; IPR001026; ENTH.
 DR InterPro; IPR008943; PI_bind_N.
 DR Pfam; PF01417; ENTH; 1.
 DR SMART; SM00273; ENTH; 1.
 DR PROSITE; PS50942; ENTH; 1.
 KW Coated pits; Alternative splicing; Phosphorylation.
 FT DOMAIN 14 145 ENTH.
 FT DOMAIN 410 413 POLY-THR.
 FT DOMAIN 535 539 POLY-ALA.
 FT DOMAIN 547 550 POLY-ALA.
 FT DOMAIN 678 683 POLY-SER.
 FT DOMAIN 723 729 POLY-SER.
 FT VARSPPLIC 614 632 Missing (in isoform Short).
 FT
 SQ SEQUENCE 915 AA; 93518 MW; 32EC1B38C5DF8C0 CRC64;
 Query Match 25.1%; Score 166.5; DB 1; Length 915;
 Best Local Similarity 37.0%; Pred. No. 0.015;
 Matches 60; Conservative 9; Mismatches 29; Gaps 6;
 QY 9 APISVP-APVAAAATAATATATATATTTTAAAPVAV-----AAAAAPAAAAAPS----- 58

Db 396 APISDFFAPEPSPPTTTTTPASASATTAATAATTEVDLEGDAPASPGEPASSEGAT 455
 QY 59 -PATAAATAAAVSPAAGQIPAAASVASAAV-----APSAASAAAVQVAPAAPVPA 111
 Db 456 APATAPVAAALD-ACSGNDPFPAPSEGSAAAPBLDLFAMKPPETSAPVVTPTASTAPPV 514
 QY 112 PALVVPAP-----AAQASAPACTOAPTSPAPAVAP 142
 Db 515 PATAPSPAPTAVATAATTTAAATTTTATTTTAAATTTTAAATTTAAAP 556
 RESULT 13
 CU38 LOCKI
 ID CU38 LOCKI STANDARD; PRT; 163 AA.
 AC P04375;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cuticle protein 38 (LM-38) (LM-ACP 38).
 DB Cuticle protein 38 (LM-38) (LM-ACP 38).
 OC Locusta migratoria (Migratory locust).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha;
 OC Acridoidea; Acrididae; Oedipodinae; Locusta.
 OX NCBI_TaxID=7004;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=87075646; PubMed=3790088;
 RA Hoeirup P., Andersen S.O., Roepstorff P.;
 RT "Primary structure of a structural protein from the cuticle of the
 migratory locust, Locusta migratoria.";
 RL Biochem. J. 236:713-720 (1986).
 RN [2]
 RP SEQUENCE OF 1-30.
 RX MEDLINE=86108304; PubMed=3943519;
 RA Hoeirup P., Andersen S.O., Roepstorff P.;
 RT "Isolation, characterization, and N-terminal sequence studies of
 cuticular proteins from the migratory locust, Locusta migratoria.";
 RL Eur. J. Biochem. 154:153-159 (1986).
 CC -!- FUNCTION: Component of the cuticle of migratory locust which
 contains more than 100 different structural proteins.
 CC -!- DOMAIN: The tetrapeptide (A-A-P-[AV]) repeats found throughout the
 protein are also present in many proteins constituting the
 protective envelope of other species.
 CC -!- SIMILARITY: TO LM-8, LM-67 AND LM-70.
 DR FIR; A03327; UCLQ38.
 KW Structural protein; Cuticle; Repeat.
 FT REPEAT 7 10 1.
 FT REPEAT 13 16 2.
 FT REPEAT 20 23 3.
 FT REPEAT 26 29 4.
 FT REPEAT 32 35 5.
 FT REPEAT 38 41 6.
 FT REPEAT 44 47 7.
 FT REPEAT 50 53 8.
 FT REPEAT 56 59 9.
 FT REPEAT 62 65 10.
 FT REPEAT 68 71 11.
 FT REPEAT 74 77 12.
 FT REPEAT 80 83 13.
 FT REPEAT 86 89 14.
 FT REPEAT 92 95 15.
 FT REPEAT 98 101 16.
 FT REPEAT 104 107 17.
 FT REPEAT 110 113 18.
 FT REPEAT 116 119 19.
 FT REPEAT 122 125 20.
 FT REPEAT 128 131 21.
 FT REPEAT 134 137 22.
 FT REPEAT 140 143 23.
 FT REPEAT 146 149 24.
 FT REPEAT 152 155 25.
 FT REPEAT 158 161 26.
 FT REPEAT 164 167 27.
 FT REPEAT 170 173 28.
 FT REPEAT 176 179 29.
 FT REPEAT 182 185 30.
 FT REPEAT 188 191 31.
 FT REPEAT 194 197 32.
 FT REPEAT 200 203 33.
 FT REPEAT 206 209 34.
 FT REPEAT 212 215 35.
 FT REPEAT 218 221 36.
 FT REPEAT 224 227 37.
 FT REPEAT 230 233 38.
 FT REPEAT 236 239 39.
 FT REPEAT 242 245 40.
 FT REPEAT 248 251 41.
 FT REPEAT 254 257 42.
 FT REPEAT 260 263 43.
 FT REPEAT 266 269 44.
 FT REPEAT 272 275 45.
 FT REPEAT 278 281 46.
 FT REPEAT 284 287 47.
 FT REPEAT 290 293 48.
 FT REPEAT 296 299 49.
 FT REPEAT 302 305 50.
 FT REPEAT 308 311 51.
 FT REPEAT 314 317 52.
 FT REPEAT 320 323 53.
 FT REPEAT 326 329 54.
 FT REPEAT 332 335 55.
 FT REPEAT 338 341 56.
 FT REPEAT 344 347 57.
 FT REPEAT 350 353 58.
 FT REPEAT 356 359 59.
 FT REPEAT 362 365 60.
 FT REPEAT 368 371 61.
 FT REPEAT 374 377 62.
 FT REPEAT 380 383 63.
 FT REPEAT 386 389 64.
 FT REPEAT 392 395 65.
 FT REPEAT 398 401 66.
 FT REPEAT 404 407 67.
 FT REPEAT 410 413 68.
 FT REPEAT 416 419 69.
 FT REPEAT 422 425 70.
 FT REPEAT 428 431 71.
 FT REPEAT 434 437 72.
 FT REPEAT 440 443 73.
 FT REPEAT 446 449 74.
 FT REPEAT 452 455 75.
 FT REPEAT 458 461 76.
 FT REPEAT 464 467 77.
 FT REPEAT 470 473 78.
 FT REPEAT 476 479 79.
 FT REPEAT 482 485 80.
 FT REPEAT 488 491 81.
 FT REPEAT 494 497 82.
 FT REPEAT 500 503 83.
 FT REPEAT 506 509 84.
 FT REPEAT 512 515 85.
 FT REPEAT 518 521 86.
 FT REPEAT 524 527 87.
 FT REPEAT 530 533 88.
 FT REPEAT 536 539 89.
 FT REPEAT 542 545 90.
 FT REPEAT 548 551 91.
 FT REPEAT 554 557 92.
 FT REPEAT 560 563 93.
 FT REPEAT 566 569 94.
 FT REPEAT 572 575 95.
 FT REPEAT 578 581 96.
 FT REPEAT 584 587 97.
 FT REPEAT 590 593 98.
 FT REPEAT 596 599 99.
 FT REPEAT 602 605 100.
 FT REPEAT 608 611 101.
 FT REPEAT 614 617 102.
 FT REPEAT 620 623 103.
 FT REPEAT 626 629 104.
 FT REPEAT 632 635 105.
 FT REPEAT 638 641 106.
 FT REPEAT 644 647 107.
 FT REPEAT 650 653 108.
 FT REPEAT 656 659 109.
 FT REPEAT 662 665 110.
 FT REPEAT 668 671 111.
 FT REPEAT 674 677 112.
 FT REPEAT 680 683 113.
 FT REPEAT 686 689 114.
 FT REPEAT 692 695 115.
 FT REPEAT 698 701 116.
 FT REPEAT 704 707 117.
 FT REPEAT 710 713 118.
 FT REPEAT 716 719 119.
 FT REPEAT 722 725 120.
 FT REPEAT 728 731 121.
 FT REPEAT 734 737 122.
 FT REPEAT 740 743 123.
 FT REPEAT 746 749 124.
 FT REPEAT 752 755 125.
 FT REPEAT 758 761 126.
 FT REPEAT 764 767 127.
 FT REPEAT 770 773 128.
 FT REPEAT 776 779 129.
 FT REPEAT 782 785 130.
 FT REPEAT 788 791 131.
 FT REPEAT 794 797 132.
 FT REPEAT 800 803 133.
 FT REPEAT 806 809 134.
 FT REPEAT 812 815 135.
 FT REPEAT 818 821 136.
 FT REPEAT 824 827 137.
 FT REPEAT 830 833 138.
 FT REPEAT 836 839 139.
 FT REPEAT 842 845 140.
 FT REPEAT 848 851 141.
 FT REPEAT 854 857 142.
 FT REPEAT 860 863 143.
 FT REPEAT 866 869 144.
 FT REPEAT 872 875 145.
 FT REPEAT 878 881 146.
 FT REPEAT 884 887 147.
 FT REPEAT 890 893 148.
 FT REPEAT 896 899 149.
 FT REPEAT 902 905 150.
 FT REPEAT 908 911 151.
 FT REPEAT 914 917 152.
 FT REPEAT 920 923 153.
 FT REPEAT 926 929 154.
 FT REPEAT 932 935 155.
 FT REPEAT 938 941 156.
 FT REPEAT 944 947 157.
 FT REPEAT 950 953 158.
 FT REPEAT 956 959 159.
 FT REPEAT 962 965 160.
 FT REPEAT 968 971 161.
 FT REPEAT 974 977 162.
 FT REPEAT 980 983 163.
 FT REPEAT 986 989 164.
 FT REPEAT 992 995 165.
 FT REPEAT 998 1001 166.
 FT REPEAT 1004 1007 167.
 FT REPEAT 1010 1013 168.
 FT REPEAT 1016 1019 169.
 FT REPEAT 1022 1025 170.
 FT REPEAT 1028 1031 171.
 FT REPEAT 1034 1037 172.
 FT REPEAT 1040 1043 173.
 FT REPEAT 1046 1049 174.
 FT REPEAT 1052 1055 175.
 FT REPEAT 1058 1061 176.
 FT REPEAT 1064 1067 177.
 FT REPEAT 1070 1073 178.
 FT REPEAT 1076 1079 179.
 FT REPEAT 1082 1085 180.
 FT REPEAT 1088 1091 181.
 FT REPEAT 1094 1097 182.
 FT REPEAT 1100 1103 183.
 FT REPEAT 1106 1109 184.
 FT REPEAT 1112 1115 185.
 FT REPEAT 1118 1121 186.
 FT REPEAT 1124 1127 187.
 FT REPEAT 1130 1133 188.
 FT REPEAT 1136 1139 189.
 FT REPEAT 1142 1145 190.
 FT REPEAT 1148 1151 191.
 FT REPEAT 1154 1157 192.
 FT REPEAT 1160 1163 193.
 FT REPEAT 1166 1169 194.
 FT REPEAT 1172 1175 195.
 FT REPEAT 1178 1181 196.
 FT REPEAT 1184 1187 197.
 FT REPEAT 1190 1193 198.
 FT REPEAT 1196 1199 199.
 FT REPEAT 1202 1205 200.
 FT REPEAT 1208 1211 201.
 FT REPEAT 1214 1217 202.
 FT REPEAT 1220 1223 203.
 FT REPEAT 1226 1229 204.
 FT REPEAT 1232 1235 205.
 FT REPEAT 1238 1241 206.
 FT REPEAT 1244 1247 207.
 FT REPEAT 1250 1253 208.
 FT REPEAT 1256 1259 209.
 FT REPEAT 1262 1265 210.
 FT REPEAT 1268 1271 211.
 FT REPEAT 1274 1277 212.
 FT REPEAT 1280 1283 213.
 FT REPEAT 1286 1289 214.
 FT REPEAT 1292 1295 215.
 FT REPEAT 1298 1301 216.
 FT REPEAT 1304 1307 217.
 FT REPEAT 1310 1313 218.
 FT REPEAT 1316 1319 219.
 FT REPEAT 1322 1325 220.
 FT REPEAT 1328 1331 221.
 FT REPEAT 1334 1337 222.
 FT REPEAT 1340 1343 223.
 FT REPEAT 1346 1349 224.
 FT REPEAT 1352 1355 225.
 FT REPEAT 1358 1361 226.
 FT REPEAT 1364 1367 227.
 FT REPEAT 1370 1373 228.
 FT REPEAT 1376 1379 229.
 FT REPEAT 1382 1385 230.
 FT REPEAT 1388 1391 231.
 FT REPEAT 1394 1397 232.
 FT REPEAT 1400 1403 233.
 FT REPEAT 1406 1409 234.
 FT REPEAT 1412 1415 235.
 FT REPEAT 1418 1421 236.
 FT REPEAT 1424 1427 237.
 FT REPEAT 1430 1433 238.
 FT REPEAT 1436 1439 239.
 FT REPEAT 1442 1445 240.
 FT REPEAT 1448 1451 241.
 FT REPEAT 1454 1457 242.
 FT REPEAT 1460 1463 243.
 FT REPEAT 1466 1469 244.
 FT REPEAT 1472 1475 245.
 FT REPEAT 1478 1481 246.
 FT REPEAT 1484 1487 247.
 FT REPEAT 1490 1493 248.
 FT REPEAT 1496 1499 249.
 FT REPEAT 1502 1505 250.
 FT REPEAT 1508 1511 251.
 FT REPEAT 1514 1517 252.
 FT REPEAT 1520 1523 253.
 FT REPEAT 1526 1529 254.
 FT REPEAT 1532 1535 255.
 FT REPEAT 1538 1541 256.
 FT REPEAT 1544 1547 257.
 FT REPEAT 1550 1553 258.
 FT REPEAT 1556 1559 259.
 FT REPEAT 1562 1565 260.
 FT REPEAT 1568 1571 261.
 FT REPEAT 1574 1577 262.
 FT REPEAT 1580 1583 263.
 FT REPEAT 1586 1589 264.
 FT REPEAT 1592 1595 265.
 FT REPEAT 1598 1601 266.
 FT REPEAT 1604 1607 267.
 FT REPEAT 1610 1613 268.
 FT REPEAT 1616 1619 269.
 FT REPEAT 1622 1625 270.
 FT REPEAT 1628 1631 271.
 FT REPEAT 1634 1637 272.
 FT REPEAT 1640 1643 273.
 FT REPEAT 1646 1649 274.
 FT REPEAT 1652 1655 275.
 FT REPEAT 1658 1661 276.
 FT REPEAT 1664 1667 277.
 FT REPEAT 1670 1673 278.
 FT REPEAT 1676 1679 279.
 FT REPEAT 1682 1685 280.
 FT REPEAT 1688 1691 281.
 FT REPEAT 1694 1697 282.
 FT REPEAT 1700 1703 283.
 FT REPEAT 1706 1709 284.
 FT REPEAT 1712 1715 285.
 FT REPEAT 1718 1721 286.
 FT REPEAT 1724 1727 287.
 FT REPEAT 1730 1733 288.
 FT REPEAT 1736 1739 289.
 FT REPEAT 1742 1745 290.
 FT REPEAT 1748 1751 291.
 FT REPEAT 1754 1757 292.
 FT REPEAT 1760 1763 293.
 FT REPEAT 1766 1769 294.
 FT REPEAT 1772 1775 295.
 FT REPEAT 1778 1781 296.
 FT REPEAT 1784 1787 297.
 FT REPEAT 1790 1793 298.
 FT REPEAT 1796 1799 299.
 FT REPEAT 1802 1805 300.
 FT REPEAT 1808 1811 301.
 FT REPEAT 1814 1817 302.
 FT REPEAT 1820 1823 303.
 FT REPEAT 1826 1829 304.
 FT REPEAT 1832 1835 305.
 FT REPEAT 1838 1841 306.
 FT REPEAT 1844 1847 307.
 FT REPEAT 1850 1853 308.
 FT REPEAT 1856 1859 309.
 FT REPEAT 1862 1865 310.
 FT REPEAT 1868 1871 311.
 FT REPEAT 1874 1877 312.
 FT REPEAT 1880 1883 313.
 FT REPEAT 1886 1889 314.
 FT REPEAT 1892 1895 315.
 FT REPEAT 1898 1901 316.
 FT REPEAT 1904 1907 317.
 FT REPEAT 1910 1913 318.
 FT REPEAT 1916 1919 319.
 FT REPEAT 1922 1925 320.
 FT REPEAT 1928 1931 321.
 FT REPEAT 1934 1937 322.
 FT REPEAT 1940 1943 323.
 FT REPEAT 1946 1949 324.
 FT REPEAT 1952 1955 325.
 FT REPEAT 1958 1961 326.
 FT REPEAT 1964 1967 327.
 FT REPEAT 1970 1973 328.
 FT REPEAT 1976 1979 329.
 FT REPEAT 1982 1985 330.
 FT REPEAT 1988 1991 331.
 FT REPEAT 1994 1997 332.
 FT REPEAT 2000 2003 333.
 FT REPEAT 2006 2009 334.
 FT REPEAT 2012 2015 335.
 FT REPEAT 2018 2021 336.
 FT REPEAT 2024 2027 337.
 FT REPEAT 2030 2033 338.
 FT REPEAT 2036 2039 339.
 FT REPEAT 2042 2045 340.
 FT REPEAT 2048 2051 341.
 FT REPEAT 2054 2057 342.
 FT REPEAT 2060 2063 343.
 FT REPEAT 2066 2069 344.
 FT REPEAT 2072 2075 345.
 FT REPEAT 2078 2081 346.
 FT REPEAT 2084 2087 347.
 FT REPEAT 2090 2093 348.
 FT REPEAT 2096 2099 349.
 FT REPEAT 2102 2105 350.
 FT REPEAT 2108 2111 351.
 FT REPEAT 2114 2117 352.
 FT REPEAT 2120 2123 353.
 FT REPEAT 2126 2129 354.
 FT REPEAT 2132 2135 355.
 FT REPEAT 2138 2141 356.
 FT REPEAT 2144 2147 357.
 FT REPEAT 2150 2153 358.
 FT REPEAT 2156 2159 359.
 FT REPEAT 2162 2165 360.
 FT REPEAT 2168 2171 361.
 FT REPEAT 2174 2177 362.
 FT REPEAT 2180 2183 363.
 FT REPEAT 2186 2189 364.
 FT REPEAT 2192 2195 365.
 FT REPEAT 2198 2201 366.
 FT REPEAT 2204 2207 367.
 FT REPEAT 2210 2213 368.
 FT REPEAT 2216 2219 369.
 FT REPEAT 2222 2225 370.
 FT REPEAT 2228 2231 371.
 FT REPEAT 2234 2237 372.
 FT REPEAT 2240 2243 373.
 FT REPEAT 2246 2249 374.
 FT REPEAT 2252 2255 375.
 FT REPEAT 2258 2261 376.
 FT REPEAT 2264 2267 377.
 FT REPEAT 2270 2273 378.
 FT REPEAT 2276 2279 379.
 FT REPEAT 2282 2285 380.
 FT REPEAT 2288 2291 381.
 FT REPEAT 2294 2297 382.
 FT REPEAT 2300 2303 383.
 FT REPEAT 2306 2309 384.
 FT REPEAT 2312 2315 385.
 FT REPEAT 2318 2321 386.
 FT REPEAT 2324 2327 387.
 FT REPEAT 2330 2333 388.
 FT REPEAT 2336 2339 389.
 FT REPEAT 2342 2345 390.
 FT REPEAT 2348 2351 391.
 FT REPEAT 2354 2357 392.
 FT REPEAT 2360 2363 393.
 FT REPEAT 2366 2369 394.
 FT REPEAT 2372 2375 395.
 FT REPEAT 2378 2381 396.
 FT REPEAT 2384 2387 397.
 FT REPEAT 2390 2393 398.
 FT REPEAT 2396 2399 399.
 FT REPEAT 2402 2405 400.
 FT REPEAT 2408 2411 401.
 FT REPEAT 2414 2417 402.
 FT REPEAT 2420 2423 403.
 FT REPEAT 2426 2429 404.
 FT REPEAT 2432 2435 405.
 FT REPEAT 2438 2441 406.
 FT REPEAT 2444 2447 407.
 FT REPEAT 2450 2453 408.
 FT REPEAT 2456 2459 409.
 FT REPEAT 2462 2465 410.
 FT REPEAT 2468 2471 411.
 FT REPEAT 2474 2477 412.
 FT REPEAT 2480 2483 413.
 FT REPEAT 2486 2489 414.
 FT REPEAT 2492 2495 415.
 FT REPEAT 2498 2501 416.
 FT REPEAT 2504 2507 417.
 FT REPEAT 2510 2513 418.
 FT REPEAT 2516 2519 419.
 FT REPEAT 2522 2525 420.
 FT REPEAT 2528 2531 421.
 FT REPEAT 2534 2537 422.
 FT REPEAT 2540 2543 423.
 FT REPEAT 2546 2549 424.
 FT REPEAT 2552 2555 425.
 FT REPEAT 2558 2561 426.
 FT REPEAT 2564 2567 427.
 FT REPEAT 2570 2573 428.
 FT REPEAT 2576 2579 429.
 FT REPEAT 2582 2585 430.
 FT REPEAT 2588 2591 431.
 FT REPEAT 2594 2597 432.
 FT REPEAT 2600 2603 433.
 FT REPEAT 2606 2609 434.
 FT REPEAT 2612 2615 435.
 FT REPEAT 2618 2621 436.
 FT REPEAT 2624 2627 437.
 FT REPEAT 2630 2633 438.
 FT REPEAT 2636 2639 439.
 FT REPEAT 2642 2645 440.
 FT REPEAT 2648 2651 441.
 FT REPEAT 2654 2657 442.
 FT REPEAT 2660 2663 443.
 FT REPEAT 2666 2669 444.
 FT REPEAT 2672 2675 445.
 FT REPEAT 2678 2681 446.
 FT REPEAT 2684 2687 447.
 FT REPEAT 2690 2693 448.
 FT REPEAT 2696 2699 449.
 FT REPEAT 2702 2705 450.
 FT REPEAT 2708 2711 451.
 FT REPEAT 2714 2717 452.
 FT REPEAT 2720 2723 453.
 FT REPEAT 2726 2729 454.
 FT REPEAT 2732 2735 455.
 FT REPEAT 2738 2741 456.
 FT REPEAT 2744 2747 457.
 FT REPEAT 2750 2753 458.
 FT REPEAT 2756 2759 459.
 FT REPEAT 2762 2765 460.
 FT REPEAT 2768 2771 461.
 FT REPEAT 2774 2777 462.
 FT REPEAT 2780 2783 463.
 FT REPEAT 2786 2789 464.
 FT REPEAT 2792 2795 465.
 FT REPEAT 2798 2801 466.
 FT REPEAT 2804 2807 467.
 FT REPEAT 2810 2813 468.
 FT REPEAT 2816 2819 46

```

Db      59  AYAAPAAVAAPAIQYAAAPATAAAPALGARYAAAAFPV-AVAHAAPAAASVANTYRISQ 117
Qy      131  TOAPTSAPAVA 141
       1  :|||||
Db      118  TARLLAAPAVA 128

RESULT 14
ID      HXAD_MOUSE
AC      Q62424;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Homeobox protein Hox-A13 (Hox-1.10).
GN      HOXA13 OR HOX-1.10.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Theria; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI    TaxID=10090;
RN      [1]_
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96259555; PubMed=8673126;
RA      Mortlock D.P., Post L.C., Innis J.W.;
RT      "The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads
RT      to arrest of digital arch formation.";
RL      Nat. Genet. 13:284-289(1995).
CC      -!- FUNCTION: Sequence-specific transcription factor which is part of
CC      a developmental regulatory system that provides cells with
CC      specific positional identities on the anterior-posterior axis (by
CC      similarity).
CC      -!- DISEASE: Defects in HOXA13 are the cause of hypodactyly (Hd), a
CC      condition characterized by profound deficiency of digital arch
CC      structures.
CC      -!- SIMILARITY: Belongs to the Abd-B homeobox family.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sb-sib.ch).
CC
JR      EMBL; U59322; AAB03322.1; -
JR      HSSP; P14653; 1B72
JR      TRANSFAC; T03337; -
JR      MGD; MGI:96173; Hoxa13.
JR      InterPro; IPR001356; Homeobox.
JR      Pfam; PF00046; homeobox; 1.
JR      ProDom; PD000010; Homeobox; 1.
JR      SMART; SM00389; HOX; 1.
JR      PROSITE; PS00027; HOMEBOX 1; 1.
JR      PROSITE; PS00071; HOMEBOX 2; 1.
JR      Homeobox; DNA-binding; Developmental protein; Nuclear protein;
JR      Transcription regulation.
JR      DOMAIN 38 51 POLY-ALA.
JR      DNA BIND 320 379 HOMEBOX.
JR      DOMAIN 52 57 POLY-GLY.
JR      DOMAIN 62 66 POLY-ALA.
JR      DOMAIN 73 84 POLY-ALA.
JR      DOMAIN 101 104 POLY-ALA.
JR      DOMAIN 116 133 POLY-ALA.
JR      DOMAIN 198 205 POLY-ALA.
JR      SEQUENCE 386 AA; 39566 MW; 2501DCC9B1951324 CRC64;

Query Match 25.0%; Score 166; DB 1; Length 386;
Best Local Similarity 40.0%; Pred. No. 0.0084;
Matches 56; Conservative 13; Mismatches 50; Indels 22; Gaps 4;

Yr      18  AAAATAAAATATATATI-----TTTAAAPVAAAAAPAAAAAPS-----PATA 62
       1  :|||||

```

```
DR Pfam: PF03144; GTP EFTU D2; 1.
DR Pfam: PF04760; IF2_N; 2.
DR PRINTS: PR003115; ELONGATNCT.
DR PRINTS: PR012117; PRICHEXTNSN.
DR PRINTS: PR004449; RASTRNSFRNG.
DR ProDom: PD186100; IF2; 1.
DR SMART: SM00173; RAS; 1.
DR TIGRFAMs: TIGR00487; IF-2; 1.
DR TIGRFAMs: TIGR00231; small GTP; 1.
DR PROSITE: PS01176; IF2; FALSE NEG.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 542 694 G-DOMAIN.
FT NP_BIND 548 555 GTP (BY SIMILARITY).
FT NP_BIND 598 602 GTP (BY SIMILARITY).
FT NP_BIND 652 655 GTP (BY SIMILARITY).
SQ SEQUENCE 1046 AA; 106882 MW; 62B12E2DFA3AA595 CRC64;

Query Match 25.0%; Score 165.5; DB 1; Length 1046;
Best Local Similarity 43.3%; Pred. No. 0.019;
Matches 58; Conservative 4; Mismatches 47; Indels 25; Gaps 7;

2Y 9 APISVPAPVAAAATAAATATAATITTTWVAAAAPVAVAAAAAPAAAAAPSPATAAATAA 68
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 ASSTIEAPVVKLTDLQCGNGG-----KAAP-----RKAAPAKPCAPSPQAARPAAP 82
2Y 69 VSPAAGCIPAAASVASAAVAPSAASAAAAAVQVAPAPAPVPAPALVVPAPAAAAQASAP 128
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
83 RPPAPK---PAAAE-RPAAERPA-----APAAPGPRPGPKAPRPAPAPAPAPAP 130
2Y 129 AQTOAPTSAPAVAP 142
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
131 EFT-APSPAP-AP 142
```

Search completed: June 16, 2004, 19:17:45
Job time : 4.49633 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - protein search, using sw model

run on: June 16, 2004, 19:17:48 ; Search time 51 seconds
(without alignments)

4175.976 Million cell updates/sec

title: US-09-332-063-2

perfect score: 675

sequence: 1 MPRAQSSASQVPADPEA.....KTPQILGQEPDAEMVEVLI 675

scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

searched: 1017041 seqs, 315518202 residues

ord size : 0

total number of hits satisfying chosen parameters: 1017041

linimum DB seq length: 0

maximum DB seq length: 2000000000

ost-processing: Listing first 45 summaries

atabase :

- 1: sp_arChaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB ID	Description
1	675	100.0	675	4 Q9HD27	Q9hd27 homo sapien
2	103	15.3	891	11 Q8VHG2	Q8vhg2 mus musculus
3	24	3.6	326	4 Q8NDNO	Q8ndno homo sapien
4	24	3.6	498	4 Q96CM5	Q96cm5 homo sapien
5	24	3.6	882	11 Q9D4H4	Q9d4h4 mus musculus
6	24	3.6	956	4 Q8IY63	Q8iy63 homo sapien
7	18	2.7	402	4 Q8TBR8	Q8tbr8 homo sapien
8	18	2.7	463	11 Q8BP84	Q8bp84 mus musculus
9	18	2.7	463	11 Q8BS08	Q8bs08 mus musculus
10	18	2.7	463	11 Q7TPE4	Q7tpe4 mus musculus
11	18	2.7	466	4 Q96F99	Q96f99 homo sapien
12	18	2.7	569	4 Q9UKB4	Q9ukb4 homo sapien
13	18	2.7	586	11 Q9QUS0	Q9qus0 mus musculus
14	18	2.7	772	11 Q8K371	Q8k371 mus musculus
15	18	2.7	859	4 Q9Y2J4	Q9y2j4 homo sapien
16	13	1.9	557	4 Q8WDX1	Q8wdx1 homo sapien

17 13 1.9 612 4 Q8TEN8
18 12 1.8 172 11 Q60586
19 11 1.6 130 2 Q8VUP9
20 11 1.6 170 16 Q7H6G5
21 11 1.6 174 16 Q7MM27
22 11 1.6 203 11 Q9Q2P5
23 11 1.6 270 10 Q7XWM7
24 11 1.6 270 16 Q8A3H3
25 11 1.6 315 5 Q8WQJ2
26 11 1.6 407 16 Q8BFB0
27 11 1.6 510 16 Q8P5B3
28 11 1.6 1604 10 Q7XQ14
29 10 1.5 82 13 Q9YH51
30 10 1.5 104 17 Q8TI79
31 10 1.5 104 17 Q8PY50
32 10 1.5 121 16 Q8EVJ1
33 10 1.5 128 16 Q8FSA0
34 10 1.5 136 4 Q8GQ14
35 10 1.5 149 16 Q7WFP90
36 10 1.5 149 16 Q7W3W0
37 10 1.5 149 16 Q7VUS8
38 10 1.5 185 5 Q18507
39 10 1.5 198 16 Q9A8H3
40 10 1.5 203 16 Q8PL97
41 10 1.5 308 2 Q93QT0
42 10 1.5 309 2 Q93QT1
43 10 1.5 314 5 Q963B8
44 10 1.5 314 5 Q8GGS6
45 10 1.5 315 5 Q8MQT0

ALIGNMENTS

RESULT 1

Q9HD27
ID Q9HD27 PRELIMINARY; PRT; 675 AA.
AC Q9HD27; Q9UPT1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Angiomotin (Hypothetical protein KIAA1071).
GN KIAA1071.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21157403; PubMed=11257124;
RA Troyanovsky B., Levchenko T., Mansson G., Matvijenko O., Holmgren L.;
RT "Angiomotin. An angiotensin binding protein that regulates endothelial cell migration and tube formation.";
RL J. Cell Biol. 152:1247-1254(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hirotsawa M., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
DR EMBL; AF286598; AAG01851.1; -;
DR EMBL; AB028994; BAA83023.3; -;
KW Hypothetical protein.
SQ SEQUENCE 675 AA; 72540 MW; EBC28B74427AD481 CRC64;

Query Match 100.0%; Score 675; DB 4; Length 675;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPRAQPSASQPVADPFAIVSRAQQMVEILSDENRNRLRQELGGCYEKVARLQKVETEI 60
 DB 1 MPRAQPSASQPVADPFAIVSRAQQMVEILSDENRNRLRQELGGCYEKVARLQKVETEI 60
 QY 61 QVSEAYENLVKSSKRALEKAVNKLGEIRBWHDFNDRLEETANKOLAEKEYEG 120
 DB 61 QVSEAYENLVKSSKRALEKAVNKLGEIRBWHDFNDRLEETANKOLAEKEYEG 120
 QY 121 SEDTRKTSIQFPAKNKESOREKELEAELATARSTNEDORRHIEIRDOALSNAQAQVVKL 180
 DB 121 SEDTRKTSIQFPAKNKESOREKELEAELATARSTNEDORRHIEIRDOALSNAQAQVVKL 180
 QY 181 EELKKKQVYVDKVKMCOALVOLQACEKEBQLEHLRLTRELRELESIRIQORGNCP 240
 DB 181 EELKKKQVYVDKVKMCOALVOLQACEKEBQLEHLRLTRELRELESIRIQORGNCP 240
 QY 241 TNVSEYNAALMELLRKEEERILALEADMTKWEQYLEENVMRHFALDAATAAQAORDTT 300
 DB 241 TNVSEYNAALMELLRKEEERILALEADMTKWEQYLEENVMRHFALDAATAAQAORDTT 300
 QY 301 VISHSPNTSYDTALEARTOKEEERILMANKRCCLDMGRITLHAQIIEKIDAMIKVLQORS 360
 DB 301 VISHSPNTSYDTALEARTOKEEERILMANKRCCLDMGRITLHAQIIEKIDAMIKVLQORS 360
 QY 361 RKEPSKTEQLSCMPAKSLMSISNAGSGLSHSSTLTGSPIMEEKRDDKSKWSGLGLLG 420
 DB 361 RKEPSKTEQLSCMPAKSLMSISNAGSGLSHSSTLTGSPIMEEKRDDKSKWSGLGLLG 420
 QY 421 GDYRAEYVPTSPVPPTPLLSHSHKSGSDCSTQTERGTESNKTAAVAPISVPAPVAA 480
 DB 421 GDYRAEYVPTSPVPPTPLLSHSHKSGSDCSTQTERGTESNKTAAVAPISVPAPVAA 480
 QY 481 AATAAATATATATITTTWVAAPVAVAAAAAPAAAAAPSPATAATAAASVPAAGQIPA 540
 DB 481 AATAAATATATITTTWVAAPVAVAAAAAPAAAAAPSPATAATAAASVPAAGQIPA 540
 QY 541 AASVASAAAVAPSPAAAAAVQVAPAPVAPALVPVPAAPAAQASAPAQTAQTSAPA 600
 DB 541 AASVASAAAVAPSPAAAAAVQVAPAPVAPALVPVPAAPAAQASAPAQTAQTSAPA 600
 QY 601 VAPTPAPTPTPAVQAQVAPSPATGPGPHRLSIPSLTNCNPKDGTGPFVHNTLERKTPIQ 660
 DB 601 VAPTPAPTPTPAVQAQVAPSPATGPGPHRLSIPSLTNCNPKDGTGPFVHNTLERKTPIQ 660
 QY 661 ILGQEPDAEMVEYLI 675
 DB 661 ILGQEPDAEMVEYLI 675
 RESULT 2
 Q8VHG2 PRELIMINARY; PRT; 891 AA.
 AC Q8VHG2;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Angiomotin.
 GN AMOT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 DX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6CBAP2; TISSUE=Placenta;
 RA Troyanovsky B., Bratt A., Holmgren L.;
 RT "Mouse angiomotin."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF461135; AAL73436.1; --
 XR MGD; MGI:108440; Amot.
 SQ SEQUENCE 891 AA; 95111 MW; C0544542A4CD6088 CRC64;

Query Match 15.3%; Score 103; DB 11; Length 891;

Best Local Similarity 100.0%; Pred. No. 3e-83;
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 244 SEYNAALMELLRKEEERILALEADMTKWEQYLEENVMRHFALDAATAAQAORDTTVIS 303
 DB 398 SEYNAALMELLRKEEERILALEADMTKWEQYLEENVMRHFALDAATAAQAORDTTVIS 457
 QY 304 HSPNTSYDTALEARTOKEEERILMANKRCCLDMGRITLHAQI 346
 DB 458 HSPNTSYDTALEARTOKEEERILMANKRCCLDMGRITLHAQI 500
 RESULT 3
 Q8NDNO PRELIMINARY; PRT; 326 AA.
 AC Q8NDNO;
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN DKFZP434O2216.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 DX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Kohrer K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL833833; CAD38693.1; --
 KW Hypothetical protein.
 FT NON TER
 SQ SEQUENCE 326 AA; 35950 MW; SE28183735B67B56 CRC64;
 Query Match 3.6%; Score 24; DB 4; Length 326;
 Best Local Similarity 100.0%; Pred. No. 9.2e-13;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 256 REKEERILALEADMTKWEQYLEE 279
 DB 41 REKEERILALEADMTKWEQYLEE 64
 RESULT 4
 Q96CM5 PRELIMINARY; PRT; 498 AA.
 AC Q96CM5;
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE Similar to RIKEN cDNA 4932416D09 gene (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 DX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC014126; AAL14126.1; --
 FT NON TER
 SQ SEQUENCE 498 AA; 56333 MW; AD557A70AB2EDDD4 CRC64;
 Query Match 3.6%; Score 24; DB 4; Length 498;
 Best Local Similarity 100.0%; Pred. No. 1.3e-12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 256 REKEERILALEADMTKWEQYLEE 279
 DB 213 REKEERILALEADMTKWEQYLEE 236

```

RESULT 5
ID Q9D4H4 PRELIMINARY; PRT; 882 AA.
AC Q9D4H4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 4932416D09RIK protein.
EN AMOTL1 OR 4932416D09RIK.
NS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
UN [1]
IP SEQUENCE FROM N.A.
IC STRAIN=C57BL/6J; TISSUE=Testis;
IC MEDLINE=21085660; PubMed=11217851;
LA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
LA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
LA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
LA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
LA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
LA Fleischmann W., Gasterland T., Gissi C., King B., Kochiwa H.,
LA Kuehl P., Lewis S., Matsuo Y., Mikaido I., Pesole G., Quackenbush J.,
LA Schiraldi L.M., Stauber P., Suzuki R., Tomita M., Wagner L., Washio T.,
LA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
LA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
LA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
LA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
LA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
LA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
LA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,
LA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
LA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,
LA Hayashizaki Y.;
LA "Functional annotation of a full-length mouse cDNA collection.";
LA Nature 409:685-690(2001).
LR EMBL; AK016526; BAB30287.1; -.
LR MGD; MGI:1922973; Amotl1.
LR GO; GO:0005923; C:right junction; IDA
RQ SEQUENCE 882 AA; 98423 MW; 88AF83FFA56046 CRC64;

Query Match 3.6%; Score 24; DB 11; Length 882;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 256 REKERILALEADMTKWEQKYLEE 279
b 596 REKERILALEADMTKWEQKYLEE 619
|||||
|||||

RESULT 6
ID Q8IY63 PRELIMINARY; PRT; 956 AA.
AC Q8IY63;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to angiotensin like 1.
NS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
UN [1]
IP SEQUENCE FROM N.A.
IC TISSUE=Muscle;
LA Strausberg R.;
LR Submitted (SFP-2002) to the EMBL/GenBank/DBJ databases.
LR EMBL; BC037539; AAH37539.1; -.
RQ SEQUENCE 956 AA; 106574 MW; 6F47AE3A13816E95 CRC64;

Query Match 3.6%; Score 24; DB 4; Length 956;

Best Local Similarity 100.0%; Pred. No. 2.3e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 REKERILALEADMTKWEQKYLEE 279
Db 671 REKERILALEADMTKWEQKYLEE 694
|||||
|||||

RESULT 7
Q8TBR8 PRELIMINARY; PRT; 402 AA.
AC Q8TBR8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
UN [1]
IP SEQUENCE FROM N.A.
IC TISSUE=Pancreas;
LA Strausberg R.;
LR Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
LR EMBL; BC025981; AAH25981.1; -.
RQ SEQUENCE 402 AA; 44234 MW; 5AD4ADFOE8A83676 CRC64;

Query Match 2.7%; Score 18; DB 4; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ILALEADMTKWEQKYLEE 279
Db 182 ILALEADMTKWEQKYLEE 199
|||||
|||||

RESULT 8
Q8BP84 PRELIMINARY; PRT; 463 AA.
AC Q8BP84;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Leman coiled-coil protein.
GN AMOTL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
UN [1]
IP SEQUENCE FROM N.A.
IC STRAIN=C57BL/6J; TISSUE=Body;
IC MEDLINE=22354683; PubMed=12466851;
LA The FANTOM Consortium,
LA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK075535; BAC36853.1; -.
DR PIR; PT0534; PT0534.
DR MGD; MGI:1929286; Amotl2.
RQ SEQUENCE 463 AA; 51733 MW; 0394E0334ACFAL7E CRC64;

Query Match 2.7%; Score 18; DB 11; Length 463;
Best Local Similarity 100.0%; Pred. No. 3e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ILALEADMTKWEQKYLEE 279
Db 247 ILALEADMTKWEQKYLEE 264
|||||
|||||

```

RESULT 9					
ID	Q8BS08	PRELIMINARY;	PRT;	463 AA.	
AC	Q8BS08;				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Leman coiled-coil protein.				
EN	AMOTL2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;				
RC	MEDLINE=22354683; PubMed=12466851;				
RA	The FANTOM Consortium,				
RA	"Analysis of the mouse transcriptome based on functional annotation of				
RT	60,770 full-length cDNAs";				
RL	Nature 420:563-573(2002)."				
RL	EMBL; AK040912; BAC30740.1; -.				
DR	PIR; PTO534; PT0534.				
DR	MGI; MGI:1929286; Amotl2.				
SQ	SEQUENCE 463 AA; 51758 MW; 2F7BE52FBC8B1678 CRC64;				
Query Match 2.7%; Score 18; DB 11; Length 463;					
Best Local Similarity 100.0%; Pred.No. 3e-07;					
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
DY	262 ILALEADMTKWEQYLEE 279				
DB	247 ILALEADMTKWEQYLEE 264				
<hr/>					
RESULT 10					
ID	Q7TPE4	PRELIMINARY;	PRT;	463 AA.	
AC	Q7TPE4;				
DT	01-OCT-2003 (TrEMBLrel. 25, Created)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Angiomotin-like protein 2 variant 2.				
EN	AMOTL2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	Kessler R., Brunet J.-F., Van Meir E.G.;				
RT	"Murine homolog of the human LCCP gene, Variant 2.";				
RL	Submitted (AUG/1999) to the EMBL/GenBank/DBJ databases.				
RL	EMBL; AF175968; AADS56363.1; -.				
SQ	SEQUENCE 463 AA; 51778 MW; 1493D080B4FBBD70 CRC64;				
Query Match 2.7%; Score 18; DB 11; Length 463;					
Best Local Similarity 100.0%; Pred.No. 3e-07;					
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY	262 ILALEADMTKWEQYLEE 279				
DB	247 ILALEADMTKWEQYLEE 264				
<hr/>					
RESULT 11					
ID	Q96F99	PRELIMINARY;	PRT;	466 AA.	
ID	Q96F99				
AC	Q96F99;				
DT	01-DEC-2001 (TrEMBLrel. 19, Created)				

N (1)
P SEQUENCE FROM N.A.
A Kessler R., Brunet J.-F., Van Weir E.G., Variant 1."
T "Murine homolog of the human LCCP gene.
L Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
R EMBL; AF175967; AAD56362.2; -
R MGD; MGI:1929286; Amotl2.
T NON TER 1
R MGD; MGI:1929286; Amotl2.
Q SEQUENCE 586 AA; 64946 MW; 4103EBA23B7CB613 CRC64;
Query Match 2.7%; Score 18; DB 11; Length 586;
Best Local Similarity 100.0%; Pred. No. 3.6e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 262 ILALEDMTKWEQKYLEE 279
b 370 ILALEDMTKWEQKYLEE 387

RESULT 14
8K371
D Q8K371 PRELIMINARY; PRT; 772 AA.
C Q8K371;
T 01-OCT-2002 (TrEMBLrel. 22, Created)
T 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Similar to angiotensin like 2.
S AMOTL2.
N Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
P (1)
N SEQUENCE FROM N.A.
A Strausberg R.;
L Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; BC027824; AA427824.1; -
R MGD; MGI:1929286; Amotl2.
Q SEQUENCE 772 AA; 85278 MW; 9979F5F4D2A54989 CRC64;

Query Match 2.7%; Score 18; DB 11; Length 772;
Best Local Similarity 100.0%; Pred. No. 4.6e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 262 ILALEDMTKWEQKYLEE 279
b 556 ILALEDMTKWEQKYLEE 573

RESULT 15
9Y2J4
D Q9Y2J4 PRELIMINARY; PRT; 859 AA.
C Q9Y2J4;
T 01-NOV-1999 (TrEMBLrel. 12, Created)
T 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E Hypothetical protein KIAA0989 (Fragment).
N KIAA0989.
S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
P (1)
N SEQUENCE FROM N.A.
C TISSUE=Brain;
X MEDLINE=9246063; PubMed=10231032;
X Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
X Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
X "Prediction of the coding sequences of unidentified human genes. XIII.
X The complete sequences of 100 new cDNA clones from brain which code
X for large proteins in vitro."
X DNA Res. 6:63-70(1999).
X EMBL; AB023206; BAA76833.1; -.

DR Genew; HGNC:17812; AMOTL2.
DR InterPro; IPR000408; Reg chr_condens.
DR PROSITE; PS00626; RCL1_2; 1.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 859 AA; 94325 MW; 5AF7E211B4CF738D CRC64;
Query Match 2.7%; Score 18; DB 4; Length 859;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 262 ILALEDMTKWEQKYLEE 279
Db 638 ILALEDMTKWEQKYLEE 655
Search completed: June 16, 2004, 19:27:17
Job time : 54 secs

Result No.	Score	Query		DB	ID	Description
		Match	%			
1	11	1.6	76	6	5273901-11	Patent No. 5273901
2	11	1.6	76	6	5482709-10	Patent No. 5482709
3	10	1.5	127	4	US-09-107-532A-6784	Sequence No. 5482709
4	10	1.5	130	4	US-09-050-739-64	Sequence 64, Appl
5	10	1.5	156	1	US-08-074-121-5	Sequence 5, Appl
6	10	1.5	156	5	PCT-US94-06447-5	Sequence 5, Appl
7	10	1.5	170	4	US-09-252-991A-27069	Sequence 27069, A
8	10	1.5	180	6	5273901-7	Patent No. 5273901
9	10	1.5	180	6	5482709-6	Patent No. 5482709
10	10	1.5	308	3	US-08-413-974-2	Sequence 2, Appl
11	10	1.5	308	3	US-08-434-418-2	Sequence 2, Appl
12	10	1.5	308	3	US-08-433-288-2	Sequence 2, Appl
13	10	1.5	308	3	US-08-174-739A-2	Sequence 2, Appl
14	10	1.5	308	4	US-08-434-256-2	Sequence 2, Appl
15	10	1.5	375	4	US-09-903-814A-2	Sequence 2, Appl
16	9	1.3	642	4	US-09-489-039A-12434	Sequence 12434, A
17	9	1.3	16	2	US-08-308-494A-2	Sequence 2, Appl
18	9	1.3	22	6	5304631-11	Patent No. 5304631
19	9	1.3	52	4	US-09-621-976-4003	Sequence 4003, Ap
20	9	1.3	118	3	US-08-301-162-10	Sequence 10, Appl
21	9	1.3	118	4	US-09-461-240-10	Sequence 10, Appl
22	9	1.3	118	4	US-09-968-927-10	Sequence 10, Appl
23	9	1.3	130	4	US-09-540-014-6	Sequence 6, Appl
24	9	1.3	181	4	US-09-252-991A-23483	Sequence 23483, A
25	9	1.3	202	4	US-09-252-991A-28326	Sequence 28326, A
26	9	1.3	263	4	US-09-252-991A-22230	Sequence 22230, A
27	9	1.3	384	4	US-09-252-991A-24427	Sequence 24427, A

```

; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO:10:
; LENGTH: 76
5482709-10

Query Match
Best Local Similarity 1.6%; Score 11; DB 6; Length 76;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAPAAAPAAA 517
b 19 AAAPAAAPAAA 29

; RESULT 3
US-09-107-532A-6784
Sequence 6784, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariiello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: STC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6784:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...127
SEQUENCE DESCRIPTION: SEQ ID NO: 6784:
US-09-107-532A-6784

Query Match
Best Local Similarity 1.5%; Score 10; DB 4; Length 127;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 500 AAAPVAVAAA 509
|||||

```

```

Db 40 AAAPVAVAAA 49

; RESULT 4
US-09-050-739-64
; Sequence 64, Application US/09050739
; Patent No. 6641814
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDRINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/050,739
; CURRENT FILING DATE: 1998-03-30
; EARLIER APPLICATION NUMBER: 0376/97
; EARLIER FILING DATE: 1997-04-02
; EARLIER APPLICATION NUMBER: 1277/97
; EARLIER FILING DATE: 1997-11-10
; EARLIER APPLICATION NUMBER: 60/044,624
; EARLIER FILING DATE: 1997-04-18
; EARLIER APPLICATION NUMBER: 60/070,488
; EARLIER FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-050-739-64

Query Match 1.5%; Score 10; DB 4; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 500 AAAPVAVAAA 509
|||||
Db 36 AAAPVAVAAA 45

; RESULT 5
US-08-074-121-5
; Sequence 5, Application US/08074121
; Patent No. 5767362
; GENERAL INFORMATION:
; APPLICANT: Best, Elaine
; APPLICANT: Knauf, Vic C.
; TITLE OF INVENTION: Methods and Compositions for Modulating
; TITLE OF INVENTION: Lipid Content of Plant Tissues
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: US
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,121
; FILING DATE: 08-JUN-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Rae-Venter, Barbara

```

REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: 05938/043001
TELEPHONE: (415) 854-5277
TELEFAX: (415) 854-0875
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-074-121-5

Query Match 1.5%; Score 10; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 508 AAAAPAAAAA 517
| | | | |
b 60 AAAAPAAAAA 69

RESULT 6
CT-US94-06447-5
Sequence 5, Application PC/TUS9406447
GENERAL INFORMATION:
APPLICANT: Calgene, Inc.
TITLE OF INVENTION: Methods and Compositions for Modulating
TITLE OF INVENTION: Lipid Content of Plant Tissues
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Weil, Gotshal & Manges
STREET: 2882 Sand Hill Road, Suite 280
CITY: Menlo Park
STATE: California
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06447
FILING DATE: 06-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE-097/WO
TELEPHONE: (415) 926-6200
TELEFAX: (415) 854-3713
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
CT-US94-06447-5

Query Match 1.5%; Score 10; DB 5; Length 156;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 508 AAAAPAAAAA 517
| | | | |
b 60 AAAAPAAAAA 69

RESULT 7

US-09-252-991A-27069
; Sequence 27069, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27069
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27069

Query Match 1.5%; Score 10; DB 4; Length 170;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 508 AAAAPAAAAA 517
| | | | |
Db 74 AAAAPAAAAA 83

RESULT 8
5273901-7
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOOROITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 7
; LENGTH: 180
5273901-7

Query Match 1.5%; Score 10; DB 6; Length 180;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 507 AAAAPAAAAA 516
| | | | |
Db 64 AAAAPAAAAA 73

RESULT 9
5482709-6
; Patent No. 5482709
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
; ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,432


```
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 581,693
; FILING DATE: 12-SEP-1990
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1989
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO:6:
; LENGTH: 180
5482709-6

Query Match
Best Local Similarity 1.5%; Score 10; DB 6; Length 180;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAAAPAAAA 516
Db 64 AAAAAAPAAAA 73

RESULT 10
JS-08-413-974-2
; Sequence 2, Application US/08413974
; Patent No. 6180368
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglou, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; TITLE OF INVENTION: RyeGrass Pollen Allergen
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6180368ris
; STREET: 1 Liberty Place, 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,974
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,861
; FILING DATE:
; APPLICATION NUMBER: US/07/746,703
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohnschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 308 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-413-974-2
```

Query Match 1.5%; Score 10; DB 3; Length 308;

```
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 ATAAATAAAV 530
Db 281 ATAAATAAAV 290

RESULT 11
US-08-434-418-2
; Sequence 2, Application US/08434418
; Patent No. 6197313
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir et al.
; TITLE OF INVENTION: RYEGRASS POLLEN ALLERGEN
; FILE REFERENCE: IMI-051CND2
; CURRENT APPLICATION NUMBER: US/08/434,418
; CURRENT FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: 08/202,861
; PRIOR FILING DATE: 1994-25-02
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Lolium perenne
US-08-434-418-2

Query Match 1.5%; Score 10; DB 3; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 ATAAATAAAV 530
Db 281 ATAAATAAAV 290

RESULT 12
US-08-433-288-2
; Sequence 2, Application US/08433288
; Patent No. 6239269
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir et al.
; TITLE OF INVENTION: RYEGRASS POLLEN ALLERGEN
; FILE REFERENCE: IMI-051CND1
; CURRENT APPLICATION NUMBER: US/08/433,288
; CURRENT FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: 08/413,947
; PRIOR FILING DATE: 1995-03-30
; PRIOR APPLICATION NUMBER: 08/202,861
; PRIOR FILING DATE: 1994-02-25
; PRIOR APPLICATION NUMBER: 07/746,703
; PRIOR FILING DATE: 1991-08-16
; PRIOR APPLICATION NUMBER: 07/585,086
; PRIOR FILING DATE: 1990-10-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Lolium perenne
US-08-433-288-2

Query Match 1.5%; Score 10; DB 3; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 ATAAATAAAV 530
Db 281 ATAAATAAAV 290

RESULT 13
```

IS-08-174-739A-2
Sequence 2, Application US/08174739A
Patent No. 6265566
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir
APPLICANT: Knox, Robert B.
APPLICANT: Smith, Penelope
APPLICANT: Avtioglu, Asil
APPLICANT: Theerakulpisut, Piyada
APPLICANT: Hough, Terryn
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield, LLP
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/174,739A
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-051DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-174-739A-2

Query Match 1.5%; Score 10; DB 3; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 521 ATAAATAAAV 530
|||||
b 281 ATAAATAAAV 290

RESULT 14
S-08-434-256-2
Sequence 2, Application US/08434256
Patent No. 6451324
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir, Knox, Robert B., Smith, Penelope,
APPLICANT: Avtioglu, Asil, Theerakulpisut, Piyada, Hough, Terryn
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6451324-ris
STREET: 1 Liberty Place, 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/434,256
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hohenschutz, Liza D.
;; REGISTRATION NUMBER: 33,712
;; REFERENCE/DOCKET NUMBER: IMPH-0003
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215)568-3100
;; TELEFAX: (215)568-3949
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 308 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-434-256-2

Query Match 1.5%; Score 10; DB 4; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 521 ATAAATAAAV 530
|||||
Db 281 ATAAATAAAV 290

RESULT 15
US-09-303-814A-2
Sequence 2, Application US/09903814A
Patent No. 6680428
GENERAL INFORMATION:
APPLICANT: Falco, Carl
APPLICANT: Pamodu, Lavo O.
APPLICANT: Orozco, Buddy
APPLICANT: Rafalski, Antoni
APPLICANT: Thorpe, Cathy
TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
FILE REFERENCE: BB1179 USDIV
CURRENT APPLICATION NUMBER: US/09/903,814A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: US 60/092,869
PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: US 09/351,703
PRIOR FILING DATE: 1999-07-12
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 375
TYPE: PRT
ORGANISM: Zea mays
US-09-303-814A-2

Query Match 1.5%; Score 10; DB 4; Length 375;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 474 VPAPVAAAAT 483
|||||
Db 247 VPAPVAAAAT 256

Search completed: June 16, 2004, 19:31:43
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM protein - protein search, using sw model

Run on: June 16, 2004, 19:27:24 ; Search time 49 Seconds
(without alignments)
3880.921 Million cell updates/sec

title: US-09-332-063-2
effect score: 675
sequence: 1 MPRAQSSASVQVPADPFA.....KTPQILGQEPDAENVLYI 675

coring table: OLIGO
Gapop 60.0 , Gapext 60.0

searched: 1158786 seqs, 281726120 residues

ord size : 0

total number of hits satisfying chosen parameters: 1158786

minimum DB seq length: 0
maximum DB seq length: 2000000000

post-processing: Listing first 45 summaries

database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	55	8.1	208	9	US-09-864-761-36456
2	48	7.1	48	9	US-09-864-761-39389
3	26	3.9	26	9	US-09-864-761-43939
4	24	3.6	688	14	US-10-204-887-87
5	24	3.6	882	14	US-10-298-417-2
6	18	2.7	467	15	US-10-094-466-52
7	18	2.7	772	14	US-10-298-417-4
8	11	1.6	228	12	US-10-243-552-605
9	11	1.6	407	15	US-10-369-493-13871
10	11	1.6	443	15	US-10-369-493-10023
11	11	1.6	1966	15	US-10-369-493-3424
12	10	1.5	120	12	US-10-282-122A-60991
13	10	1.5	120	16	US-10-389-566-1360
14	10	1.5	120	16	US-10-389-566-1360
15	10	1.5	122	9	US-09-815-242-13247

Sequence 13588, A
Sequence 57450, A
Sequence 6, Appli
Sequence 20, Appli
Sequence 1441, Ap
Sequence 1863, Ap
Sequence 1737, Ap
Sequence 1825, Ap
Sequence 2228, Ap
Sequence 1314, Ap
Sequence 1365, Ap
Sequence 1531, Ap
Sequence 1546, Ap
Sequence 53900, A
Sequence 7415, Ap
Sequence 12426, A
Sequence 1270, Ap
Sequence 4044, Ap
Sequence 7, Appli
Sequence 8, Appli
Sequence 1051, Ap
Sequence 64, Appli
Sequence 61922, A
Sequence 62378, A
Sequence 64413, A
Sequence 64, Appli
Sequence 64, Appli
Sequence 1286, Ap
Sequence 1794, Ap

ALIGNMENTS

RESULT 1

US-09-864-761-36456
; Sequence 36456, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn. Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 36456
 LENGTH: 208
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC004827.1
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.7
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.7
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.5
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.4
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.5
 OTHER INFORMATION: EST_HUMAN HIT: AI736861.1, EVALUE 2.00e-07
 US-09-864-761-36456
 Query Match 8.1%; Score 55; DB 9; Length 208;
 Best Local Similarity 100.0%; Pred. No. 1e-39;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPRAQSSASVOPVADPFAIVSRQAQWVEILSDENRLRQELSGCYEKVARLQK 55
 DB 154 MPRAQSSASVOPVADPFAIVSRQAQWVEILSDENRLRQELSGCYEKVARLQK 208
 RESULT 2
 US-09-864-761-39389
 Sequence 39389, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharron G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aemica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 39389
 LENGTH: 48
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC004827.1
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.2
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 10
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.5
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.6
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 9
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.8
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 8
 OTHER INFORMATION: EST_HUMAN HIT: BF032018.1, EVALUE 8.00e-20
 OTHER INFORMATION: SWISSPROT HIT: Q99972, EVALUE 9.40e-01
 US-09-864-761-39389
 Query Match 7.1%; Score 48; DB 9; Length 48;
 Best Local Similarity 100.0%; Pred. No. 3.7e-34;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 56 VETETQVSEAYENLVKSSKREALEKAMRNKLRGEIRRMHDFNRDLR 103
 DB 1 VETETQVSEAYENLVKSSKREALEKAMRNKLRGEIRRMHDFNRDLR 48
 RESULT 3
 US-09-864-761-43939
 Sequence 43939, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharron G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aemica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annotax Sequence Listing Engine vers. 1.1
SEQ ID NO 43939
LENGTH: 26
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004827.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.56
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.59
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.64
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.71
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.7
IS-09-864-761-43939
Query Match 3.9%; Score 26; DB 9; Length 26;
Best Local Similarity 100.0%; Pred. No. 4.1e-15;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 650 SNTLTKPTIQLGQBPDAEMVYLI 675
B 1 SNTLTKPTIQLGQBPDAEMVYLI 26
RESULT 4
US-10-204-887-87
Sequence 87, Application US/10204887
Publication No. US20030124569A1
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: PANZER, Scott R.
APPLICANT: SPIRO, Peter A.
APPLICANT: BANVILLE, Steven C.
APPLICANT: SHAH, Purvi
APPLICANT: CEALUP, Michael S.
APPLICANT: CHENG, Simon C.
APPLICANT: CHEN, Alice
APPLICANT: D'SA, Steven A.
APPLICANT: AMSEY, Stefan
APPLICANT: DAHL, Christopher R.
APPLICANT: DAM, Tam C.
APPLICANT: DANIELS, Susan E.
APPLICANT: DUFOUR, Gerard E.
APPLICANT: FLORES, Vincent
APPLICANT: FONG, Willy T.
APPLICANT: GREENAWALT, Lila B.
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: JONES, Anissa L.
APPLICANT: LIU, Tommy P.
APPLICANT: ROSEBERRY, Ann M.
APPLICANT: ROSEN, Bruce H.
APPLICANT: RUSSO, Frank D.
APPLICANT: STOCKREHER, Theresa K.
APPLICANT: DAFFO, Abel
APPLICANT: WRIGHT, Rachel J.
APPLICANT: YAP, Pierre E.
APPLICANT: YU, Jimmy Y.
APPLICANT: BRADLEY, Diana L.
APPLICANT: BRATCHER, Shawn R.
APPLICANT: CHEN, Wensheng
APPLICANT: COHEN, Howard J.
APPLICANT: HODGSON, David M.
APPLICANT: LINCOLN, Stephen E.
TITLE OF INVENTION: SECRETORY MOLECULES
FILE REFERENCE: PT-1134.PCT
CURRENT APPLICATION NUMBER: US/10/204,887
CURRENT FILING DATE: 2002-08-21
PRIOR APPLICATION NUMBER: 60/205,324; 60/205,286
PRIOR FILING DATE: 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
2000-05-17; 2000-05-17
NUMBER OF SEQ ID NOS: 159
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 608
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:212029.1.orf2:2000FEB01
FEATURE:
NAME/KEY: unsure
LOCATION: 18, 388
OTHER INFORMATION: unknown or other
US-10-204-887-87
Query Match 3.6%; Score 24; DB 14; Length 608;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 256 REKEERILALEADMTKWEQKYLEE 279
DB 323 REKEERILALEADMTKWEQKYLEE 346
RESULT 5
US-10-298-417-2
Sequence 2, Application US/10298417
Publication No. US20030124603A1
GENERAL INFORMATION:
APPLICANT: Miyuki Nishimura
APPLICANT: Mayumi Asano
APPLICANT: Yuichi Ono
APPLICANT: Koji Morimoto
APPLICANT: Masakazu Takeuchi
APPLICANT: Yoko Inoue
APPLICANT: Toshio Imai
APPLICANT: Yoshimi Takai
TITLE OF INVENTION: Exocrine gland tight junction-constituting protein JPAP family
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/298,417
CURRENT FILING DATE: 2003-01-22
PRIOR APPLICATION NUMBER: JP 2001-352241
PRIOR FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 882
TYPE: PRT
ORGANISM: Mus musculus
US-10-298-417-2

```
Query Match
Best Local Similarity 100.0%; Score 24; DB 14; Length 882;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 REKEERILALEADMTKWEQKYLEE 279
Db 596 REKEERILALEADMTKWEQKYLEE 619

RESULT 6
US-10-094-466-52
; Sequence 52, Application US/10094466
; Publication No. US20030203363A1
; GENERAL INFORMATION:
; APPLICANT: Spyttek et al.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
; TITLE OF INVENTION: AND METHODS OF USING
; FILE REFERENCE: 21402-290D
; CURRENT APPLICATION NUMBER: US/10/094,466
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/288,148
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/338,375
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/275,579
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/335,302
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/275,601
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/276,000
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/277,338
; PRIOR FILING DATE: 2001-03-20
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatIn 2.1
; SEQ ID NO 52
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-466-52

Query Match
Best Local Similarity 100.0%; Score 18; DB 15; Length 467;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ILALEADMTKWEQKYLEE 279
Db 246 ILALEADMTKWEQKYLEE 263

RESULT 7
US-10-298-417-4
; Sequence 4, Application US/10298417
; Publication No. US20030124603A1
; GENERAL INFORMATION:
; APPLICANT: Miyuki Nishimura
; APPLICANT: Mayumi Asano
; APPLICANT: Yuichi Ono
; APPLICANT: Koji Morimoto
; APPLICANT: Masakazu Takeuchi
; APPLICANT: Yoko Inoue
; APPLICANT: Toshio Imai
; APPLICANT: Yoshimi Takai
; TITLE OF INVENTION: Exocrine gland tight junction-constituting protein JHAP family
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/298,417
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: JP 2001-352241
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 772
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-298-417-4

Query Match
Best Local Similarity 100.0%; Score 18; DB 14; Length 772;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ILALEADMTKWEQKYLEE 279
Db 556 ILALEADMTKWEQKYLEE 573

RESULT 8
US-10-243-552-605
; Sequence 605, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10/243,552
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: PC_FL_genes version 5.0
; SEQ ID NO 605
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-605

Query Match
Best Local Similarity 100.0%; Score 11; DB 12; Length 228;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAATAAAAA 517
|||||
```

78 AAAAAA 88

RESULT 9

US-10-369-493-13871
Sequence 13871, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 13871

LENGTH: 407

TYPE: PRT

ORGANISM: Pseudomonas fluorescens

US-10-369-493-13871

Query Match 1.6%; Score 11; DB 15; Length 407;

Best Local Similarity 100.0%; Pred. No. 0.73;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAAAA 517

|||||

87 AAAAAA 97

RESULT 10

US-10-369-493-10023
Sequence 10023, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 10023

LENGTH: 443

TYPE: PRT

ORGANISM: magnetite-containing magnetic coccus

US-10-369-493-10023

Query Match 1.6%; Score 11; DB 15; Length 443;

Best Local Similarity 100.0%; Pred. No. 0.78;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 602 APTAPTPTCA 612

|||||

180 APTAPTPTCA 190

RESULT 11

US-10-369-493-3424
Sequence 3424, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 3424

LENGTH: 1906

TYPE: PRT

ORGANISM: Neurospora crassa

FEATURE:

NAME/KEY: unsure

LOCATION: {1}..(1906)

OTHER INFORMATION: unsure at all Xaa locations

US-10-369-493-3424

Query Match 1.6%; Score 11; DB 15; Length 1906;

Best Local Similarity 100.0%; Pred. No. 3;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAAA 517

|||||

114 AAAAAA 124

RESULT 12

US-10-282-122A-60991
Sequence 60991, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.

;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 60991

;; LENGTH: 120

;; TYPE: PRT

;; ORGANISM: Listeria monocytogenes

US-10-282-122A-60991

Query Match 1.5%; Score 10; DB 12; Length 120;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509

|||||

Db 35 AAAPVAVAAA 44

RESULT 13

JS-10-389-566-1360

;; Sequence 1360, Application US/10389566

;; Publication No. US20040025202A1

;; GENERAL INFORMATION:

;; APPLICANT: Monsanto Technology, LLC

;; APPLICANT: Laurie, Cathy C

;; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants

;; FILE REFERENCE: 38-77(52900)D

;; CURRENT APPLICATION NUMBER: US/10/389,566

;; CURRENT FILING DATE: 2003-03-31

;; PRIOR APPLICATION NUMBER: US 60/365,301

;; PRIOR FILING DATE: 2002-03-15

;; PRIOR APPLICATION NUMBER: US 60/391,786

;; PRIOR FILING DATE: 2002-06-25

;; PRIOR APPLICATION NUMBER: US 60/392,018

;; PRIOR FILING DATE: 2002-06-26

;; NUMBER OF SEQ ID NOS: 2459

;; SOFTWARE: PatentIn version 3.2

;; SEQ ID NO 1360

;; LENGTH: 120

;; TYPE: PRT

;; ORGANISM: Listeria innocua

JS-10-389-566-1360

Query Match 1.5%; Score 10; DB 16; Length 120;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509

|||||

Db 35 AAAPVAVAAA 44

RESULT 14

JS-10-389-566-1545

;; Sequence 1545, Application US/10389566

;; Publication No. US20040025202A1

;; GENERAL INFORMATION:

;; APPLICANT: Monsanto Technology, LLC

;; APPLICANT: Laurie, Cathy C

;; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants

;; FILE REFERENCE: 38-77(52900)D

;; CURRENT APPLICATION NUMBER: US/10/389,566

;; CURRENT FILING DATE: 2003-03-31

;; PRIOR APPLICATION NUMBER: US 60/365,301

;; PRIOR FILING DATE: 2002-03-15

;; PRIOR APPLICATION NUMBER: US 60/391,786

;; PRIOR FILING DATE: 2002-06-25

;; PRIOR APPLICATION NUMBER: US 60/392,018

;; PRIOR FILING DATE: 2002-06-26

;; NUMBER OF SEQ ID NOS: 2459

;; SOFTWARE: PatentIn version 3.2

;; SEQ ID NO 1545

;; LENGTH: 120

;; TYPE: PRT

;; ORGANISM: Listeria monocytogenes

US-10-389-566-1545

Query Match 1.5%; Score 10; DB 16; Length 120;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509

|||||

Db 35 AAAPVAVAAA 44

RESULT 15

US-09-815-242-13247

;; Sequence 13247, Application US/09815242

;; Patent No. US2002061569A1

;; GENERAL INFORMATION:

;; APPLICANT: Haselbeck, Robert

;; APPLICANT: Ohlsen, Kari L.

;; APPLICANT: Zyskind, Judith W.

;; APPLICANT: Wall, Daniel

;; APPLICANT: Trawick, John D.

;; APPLICANT: Carr, Grant J.

;; APPLICANT: Yamamoto, Robert T.

;; APPLICANT: Xu, H. Howard

;; TITLE OF INVENTION: Identification of Essential Genes in

;; FILE REFERENCE: ELITRA-011A

;; CURRENT APPLICATION NUMBER: US/09/815,242

;; CURRENT FILING DATE: 2001-03-21

;; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-03-21

;; PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-05-23

;; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23

;; PRIOR APPLICATION NUMBER: 60/253,625

;; PRIOR FILING DATE: 2000-11-27

;; PRIOR APPLICATION NUMBER: 60/257,931

;; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: 60/269,308

;; PRIOR FILING DATE: 2001-02-16

;; NUMBER OF SEQ ID NOS: 14110

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 13247

;; LENGTH: 122

;; TYPE: PRT

;; ORGANISM: Streptococcus pneumoniae

US-09-815-242-13247

Query Match

1.5%; Score 10; DB 9; Length 122;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509

|||||

Db 35 AAAPVAVAAA 44

Search completed: June 16, 2004, 19:33:18

Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

% protein - protein search, using sw model
run on: June 16, 2004, 19:23:53 ; Search time 20 Seconds
(without alignments)
3246.464 Million cell updates/sec

Title: US-09-332-063-2
Effect score: 675
Sequence: 1 MPRAQSSASVQVPADPFA.....KTPQILQGPDAEMVEYLI 675

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78.*

1: Piri.*

2: Piri.*

3: Piri.*

4: Piri.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	12	1.8	172	S35568	sex-determining pr
2	11	1.6	109	R6UTP1	acidic ribosomal p
3	11	1.6	270	E87649	hypothetical prote
4	11	1.6	312	T21351	hypothetical prote
5	11	1.6	553	B55514	dihydrolipoamide S
6	10	1.5	82	F0DLAW	antifreeze protein
7	10	1.5	82	S02326	antifreeze protein
8	10	1.5	82	A05161	antifreeze protein
9	10	1.5	82	I51125	antifreeze protein
10	10	1.5	101	H69091	ribosomal protein
11	10	1.5	106	R6DQ92	acidic ribosomal p
12	10	1.5	120	AD1106	ribosomal protein
13	10	1.5	120	AD1468	ribosomal protein
14	10	1.5	122	C95157	ribosomal protein
15	10	1.5	124	I40348	ribosomal protein
16	10	1.5	124	I40350	ribosomal protein
17	10	1.5	124	AF3345	LSU ribosomal prot
18	10	1.5	125	AC2817	50S ribosomal prot
19	10	1.5	125	D97595	ribosomal protein
20	10	1.5	127	H87310	ribosomal protein
21	10	1.5	127	B98023	50S ribosomal prot
22	10	1.5	128	R7HG12	ribosomal protein
23	10	1.5	130	A70615	probable ribosomal
24	10	1.5	130	S41123	ribosomal protein
25	10	1.5	136	A56062	Alu RNA-binding pr
26	10	1.5	136	S34196	signal recognition
27	10	1.5	156	A49342	acetyl-CoA carboxy
28	10	1.5	158	R87420	hypothetical prote
29	10	1.5	308	A38582	pollen allergen pI

30 10 1.5 315 2 B99607
31 10 1.5 635 2 C82079
32 10 1.5 747 1 EABO
33 10 1.5 762 2 H87302
34 10 1.5 850 2 JC5047
35 10 1.5 1887 2 S61703
36 9 1.3 102 2 D64363
37 9 1.3 106 2 C69436
38 9 1.3 110 2 T37490
39 9 1.3 111 2 E72524
40 9 1.3 135 2 S34952
41 9 1.3 140 2 D97722
42 9 1.3 149 2 AB0770
43 9 1.3 191 2 H83151
44 9 1.3 200 2 G84822
45 9 1.3 227 2 G70555

ALIGNMENTS

RESULT 1

S35568
sex-determining protein Sry - multimammate rat (Mastomys hildebrandtii)
C:Species: Mastomys hildebrandtii
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Sep-1999
C:Accession: S35568
R:Tucker, P.K.; Lundrigan, B.L.
Nature 364, 715-717, 1993
A:Title: Rapid evolution of the sex determining locus in Old World mice and rats.
A:Reference number: S35565; MUID:93361118; PMID:8355784
A:Accession: S35568
A:Molecule type: DNA
A:Residues: 1-172 <TUC>
A:Cross-references: GB:L29542; NID:G496161; PIDN:AAA40587.1; PID:G496162
C:Genetics:
A:Gene: Sry
A:Map position: Y
C:Superfamily: unassigned HMG box proteins; HMG box homology
C:Keywords: DNA binding
P:2-77/Domain: HMG box homology <HMG1>

Query Match 1.8%; Score 12; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 507 AAAAAPAAAAAP 518
||| |||||
Db 153 AAAAAPAAAAAP 164

RESULT 2

R6UTP1
acidic ribosomal protein P1 - Trypanosoma cruzi
C:Species: Trypanosoma cruzi
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 22-Jun-1999
C:Accession: S22644; S15924
R:Vazquez, M.P.; Schijman, A.G.; Levin, M.J.
Nucleic Acids Res. 20, 2599, 1992
A:Title: Nucleotide sequence of a cDNA encoding a Trypanosoma cruzi acidic ribosomal P1
A:Reference number: S22644; MUID:92285148; PMID:1598221
A:Accession: S22644
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-109 <VAZ>
A:Cross-references: EMBL:X55025; NID:G10629; PIDN:CAA46159.1; PID:G10630
C:Superfamily: rat acidic ribosomal protein P1
C:Keywords: phosphoprotein; protein biosynthesis; ribosome

Query Match 1.6%; Score 11; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
2Y 507 AAAAAPAAAA 517
   |||||
Db 82 AAAAAPAAAA 92

RESULT 3
387649
;Protein: hypothetical protein CC3231 [imported] - Caulobacter crescentus
;Species: Caulobacter crescentus
;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
;Accession: B87649
;Nierman, W.C.; Felblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
;Title: Complete Genome Sequence of Caulobacter crescentus.
;Reference number: A87249; MUID:21173698; PMID:11259647
;Accession: B87649
;Status: preliminary
;Molecule type: DNA
;Residues: 1-270 <STO>
;Cross-references: GB:AB005673; MID:g13424915; PIDN:AAK25193.1; GSPDB:GN00148
;Genetics:
;Gene: CC3231

Query Match 1.6%; Score 11; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 602 APTAPPTPTA 612
   |||||
Db 84 APTAPPTPTA 94

RESULT 4
T21351
;Protein: hypothetical protein F25H2.10 - Caenorhabditis elegans
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
;Accession: T21351
;Wilkinson, J.
submitted to the EMBL Data Library, September 1996
;Reference number: Z19411
;Accession: T21351
;Status: preliminary; translated from GB/EMBL/DDBU
;Molecule type: DNA
;Residues: 1-312 <WIL>
;Cross-references: EMBL:Z79754; PIDN:CAB02098.1; GSPDB:GN00019; CESP:F25H2.10
;Experimental source: clone F25H2
;Genetics:
;Gene: CESP:F25H2.10
;Map position: 1
;Introns: 49/1; 157/3
;Superfamily: rat acidic ribosomal protein P0

Query Match 1.6%; Score 11; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 508 AAAAAPAAAA 518
   |||||
Db 276 AAAAAPAAAA 286

RESULT 5
35514
;Protein: dihydrolipoamide S-acetyltransferase (EC 2.3.1.12) - Alcaligenes eutrophus
;Species: Alcaligenes eutrophus
;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 05-May-2000
;Accession: B5514
;Hein, S.; Steinbuechel, A.
Bacteriol. 176, 4394-4408, 1994
;Title: Biochemical and molecular characterization of the Alcaligenes eutrophus pyruvat
```

```
A:Reference number: A55514; MUID:94292470; PMID:8021225
A:Accession: B5514
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-553 <HEI>
A:Cross-references: GB:U09865; MID:g497263; PIDN:AAA21599.1; PID:g497265
C:Genetics:
A:Gene: pdhE
C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
F:6-78/Domain: lipoyl/biotin-binding homology <LBP1>
F:124-196/Domain: lipoyl/biotin-binding homology <LBP2>
F:526,530/Active site: His, Asp #status predicted

Query Match 1.6%; Score 11; DB 2; Length 553;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 507 AAAAAPAAAA 517
   |||||
Db 199 AAAAAPAAAA 209

RESULT 6
FDLAW
;Protein: antifreeze protein A precursor - winter flounder
;Species: Pseudopleuronectes americanus (winter flounder)
;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 22-Jun-1999
;Accession: J50704; A03194
;Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:92209995; PMID:1555765
A:Accession: J50704
A:Molecule type: DNA
A:Residues: 1-82 <DAL>
A:Cross-references: GB:M62412; GB:M62416; MID:g213592; PIDN:AAA49471.1; PID:g213593
;Davies, P.L.; Roach, A.H.; Hew, C.L.
Proc. Natl. Acad. Sci. U.S.A. 79, 335-339, 1982
A:Title: DNA sequence coding for an antifreeze protein precursor from winter flounder.
A:Reference number: A03194; MUID:82197490; PMID:6952188
A:Accession: A03194
A:Molecule type: mRNA
A:Residues: 1-82 <DA2>
A:Experimental source: clones 4-2b and 2A-7c
A:Note: the authors translated the codon AGC for residue 24 as Arg
C:Genetics:
A:Introns: 19/2
C:Superfamily: antifreeze protein
C:Keywords: antifreeze; plasma; tandem repeat
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-44/Domain: propeptide #status predicted <PRO>
F:45-82/Product: antifreeze protein A #status predicted <MAT>

Query Match 1.5%; Score 10; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 507 AAAAAPAAAA 516
   |||||
Db 34 AAAAAPAAAA 43

RESULT 7
S02326
;Protein: antifreeze protein A - winter flounder
;Species: Pseudopleuronectes americanus (winter flounder)
;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 24-Oct-2000
;Accession: S02326; JH0627
;Scott, G.K.; Davies, P.L.; Kao, M.H.; Fletcher, G.L.
J. Mol. Evol. 27, 29-35, 1988
A:Title: Differential amplification of antifreeze protein genes in the Pleuronectinae.
A:Reference number: S02326; MUID:88259236; PMID:313486
```

Accession: S02326
Molecule type: DNA
Residues: 1-82 <SC0>
Cross-references: EMBL:X07506; MID:g64211; PIDN:CAA30389.1; PID:g64212
Davies, P.L.
Gene 112, 163-170, 1992
Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
Reference number: JH0627; MUID:9220995; PMID:1555765
Accession: JH0627
Molecule type: DNA
Residues: 1-82 <DAV>
Cross-references: GB:M62415
Genetics:
Introns: 19/2
Superfamily: antifreeze protein
Keywords: antifreeze

Query Match 1.5%; Score 10; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAAAAPAAAA 516
|||||
b 34 AAAAAAPAAAA 43

RESULT 8
05161
ntifreeze protein B precursor - winter flounder
Species: Pseudopleuronectes americanus (winter flounder)
Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 24-Oct-2000
Accession: A05161
Davies, P.L.; Hough, C.; Scott, G.K.; Ng, N.; White, B.N.; Hew, C.L.
Biochem. J. 259, 9241-9247, 1984
Reference number: A05161; MUID:84264559; PMID:6086629
Accession: A05161
Molecule type: DNA
Residues: 1-82 <DAV>
Cross-references: GB:L00138; GB:J00929; MID:g343126; PIDN:AAE59964.1; PID:g457351
Genetics:
Introns: 19/2
Superfamily: antifreeze protein
Keywords: antifreeze

Query Match 1.5%; Score 10; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAAAPAAAA 516
|||||
Db 34 AAAAAAPAAAA 43

RESULT 10
H69091
ribosomal protein Lp1 - Methanobacterium thermoautotrophicum (strain Delta H)
Species: Methanobacterium thermoautotrophicum
Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 13-Aug-1999
Accession: H69091
R. Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Adredge, T.;
Ou, D.; Spadafora, R.; Vicair, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Gwani, N.
Kl, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
Reference number: A69000; MUID:98037514; PMID:9371463
Accession: H69091
Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-101 <MTH>
Cross-references: GB:AE000926; GB:AE000666; MID:g2622806; PIDN:AAE86154.1; PID:g262281
Experimental source: strain Delta H
Genetics:
Gene: MTH1682
Superfamily: rat acidic ribosomal protein P1

Query Match 1.5%; Score 10; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 AAAAAAPAAAA 518
|||||
Db 60 AAAAAAPAAAA 69

RESULT 11
R6DOP2
acidic ribosomal protein P2 - slime mold (Dictyostelium discoideum)
Alternate names: ribosomal phosphoprotein P2
Species: Dictyostelium discoideum
Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 22-Jun-1999
Accession: S14014
Prieto, J.; Candel, E.; Coloma, A.
Nucleic Acids Res. 19, 1341, 1991
Title: Nucleotide sequence of a cDNA encoding acidic ribosomal phosphoprotein P2 in Di
Reference number: S14014; MUID:91232921; PMID:1840653
Accession: S14014
Status: nucleic acid sequence not shown; translation not shown
Molecule type: mRNA
Residues: 1-106 <PRI>
Cross-references: EMBL:X56192; MID:g7340; PIDN:CAA39655.1; PID:g7341
Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1990
Note: part of this sequence, including the amino end of the mature protein, was confir
Superfamily: rat acidic ribosomal protein P1
Keywords: phosphoprotein; protein biosynthesis; ribosome
F:1-106/Product: acidic ribosomal protein P2 #status experimental <MAT>
F:64-85/Region: alanine-rich

Accession: S02326
Molecule type: DNA
Residues: 1-82 <SC0>
Cross-references: EMBL:X07506; MID:g64211; PIDN:CAA30389.1; PID:g64212
Davies, P.L.
Gene 112, 163-170, 1992
Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
Reference number: JH0627; MUID:9220995; PMID:1555765
Accession: JH0627
Molecule type: DNA
Residues: 1-82 <DAV>
Cross-references: GB:M62415
Genetics:
Introns: 19/2
Superfamily: antifreeze protein
Keywords: antifreeze

Query Match 1.5%; Score 10; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAAAAPAAAA 516
|||||
b 34 AAAAAAPAAAA 43

RESULT 9
51125
ntifreeze protein - winter flounder
Species: Pseudopleuronectes americanus (winter flounder)
Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Oct-2000
Accession: I51125
Pickett, M.; Scott, G.; Davies, P.; Wang, N.; Joshi, S.; Hew, C.
ur. J. Biochem. 143, 35-38, 1984
Title: Sequence of an antifreeze protein precursor.
Reference number: I51125; MUID:84285392; PMID:6547905
Accession: I51125
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: mRNA
Residues: 1-82 <PIC>
Cross-references: GB:M28337; MID:g213581; PIDN:AAA49466.1; PID:g213582
Genetics:
Gene: AFP
Superfamily: antifreeze protein
Keywords: antifreeze

Query Match 1.5%; Score 10; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 506 VAAAAAPAAAA 515
|||||
Db 63 VAAAAAPAAAA 72

RESULT 12

```

AD1106
ribosomal protein L12 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AD1106
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1106
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00778.1; PID:gl6409616; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: rplL
C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 1.5%; Score 10; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 13
AD1468
ribosomal protein L12 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AD1468
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1468
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95516.1; PID:gl6412712; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: rplL
C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 1.5%; Score 10; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 14
95157
ribosomal protein L7/L12 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95157
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
m, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, P.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95157
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <KUP>
A:Cross-references: GB:AS005672; PIDN:AAK75452.1; PID:gl4972838; GSPDB:GN00164; TIGR:SP
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI354
C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 1.5%; Score 10; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 15
I40348
ribosomal protein L7/L12 - Brucella abortus
C:Species: Brucella abortus
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 13-Aug-1999
C:Accession: I40348
R:Oliveira, S.C.; Zhu, Y.; Splitter, G.A.
Gene 140, 137-138, 1994
A:Title: Sequences of the rplJL operon containing the L10 and L7/L12 genes from Brucella
A:Reference number: I40348; MUID:94171071; PMID:8125331
A:Accession: I40348
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-124 <RES>
A:Cross-references: GB:L19101; NID:G304270; PIDN:AAA19863.1; PID:G304271
C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 1.5%; Score 10; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZY 500 AAAPVAVAAA 509
DB 34 AAAPVAVAAA 43

Search completed: June 16, 2004, 19:27:50
Job time : 22 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: June 16, 2004, 19:17:13 ; Search time 18 Seconds
(without alignments)
1952.631 Million cell updates/sec
Title: US-09-332-063-2
Sequence: 1 MPRAQSSASVQVPADPPA.....KTPQILGQEPDAENVYLI 675
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 141681 seqs, 52070155 residues
Word size : 0
Total number of hits satisfying chosen parameters: 141681
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries
Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB ID	Description
1	11	1.6	109	1 RLAI TRYCR	P26643 trypanosoma
2	11	1.6	114	1 RLAI TRYCR	Q96799 eimeria ten
3	11	1.6	311	1 RLAI TRYCR	Q93572 caenorhabdi
4	11	1.6	553	1 ODP2_AUCEU	Q59098 alcaligenes
5	10	1.5	82	1 ANPA_PSEAM	P04002 pseudopieur
6	10	1.5	101	1 RL12 METTH	P05394 methanobact
7	10	1.5	101	1 RL12 METTH	O52706 methanococc
8	10	1.5	105	1 RL21_DICDI	P22683 dictyosteli
9	10	1.5	120	1 RL7_LISNO	Q92424 listeria in
10	10	1.5	120	1 RL7_LISNO	Q87443 listeria mo
11	10	1.5	121	1 RL7_STRPN	P80714 streptococ
12	10	1.5	122	1 RL7_FUSNN	Q8rh15 fusobacteri
13	10	1.5	124	1 RL7_BRUNE	P41106 brucella me
14	10	1.5	125	1 RL7_AGR75	Q8ue07 agrobacteri
15	10	1.5	125	1 RL7_RHIL0	Q8m67 rhizobium 1
16	10	1.5	125	1 RL7_THETH	Q8vve2 thermus the
17	10	1.5	125	1 RL7_THETH	Q8r7u5 thermomanae
18	10	1.5	126	1 RL7_RHIME	Q92488 rhizobium m
19	10	1.5	127	1 RL7_CAUCR	Q9aa07 caulobacter
20	10	1.5	128	1 RL7_CORGL	Q8nt28 corynebacte
21	10	1.5	128	1 RL7_THEMA	P29396 thermotoga
22	10	1.5	130	1 RL7_MYCTU	P37381 mycobacteri
23	10	1.5	136	1 RL7_HUMAN	P37108 homo sapien
24	10	1.5	156	1 ECCC_PSEAE	P37799 pseudomonas
25	10	1.5	157	1 ECCC_LOEMI	P45584 locusta mig
26	10	1.5	308	1 MP5A_LOLFR	Q40240 lolium pere
27	10	1.5	314	1 RLAI EIMTE	Q96797 eimeria ten
28	10	1.5	322	1 RLAI EIMTE	P50345 lupinus lut
29	10	1.5	747	1 ELS_BOVIN	P04985 bos taurus
30	10	1.5	848	1 RSG2_MOUSE	P58069 mus musculu
31	10	1.5	849	1 RSG2_HUMAN	Q15283 homo sapien
32	10	1.5	1887	1 PAS2_YEAST	P19097 s fatty aci
33	9	1.3	102	1 RL12_METTA	P54048 methanococc

ALIGNMENTS

RESULT 1
ID RLAI TRYCR STANDARD; PRT; 109 AA.
AC P26643;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 60S acidic ribosomal protein P1.
OS Trypanosoma cruzi.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92285148; PubMed=1598221;
RA Vazquez M., Schijman A., Levin M.;
RT "Nucleotide sequence of a cDNA encoding a Trypanosoma cruzi acidic ribosomal P1 type protein.";
RL Nucleic Acids Res. 20:2599-2599(1992).
CC -!- FUNCTION: Plays an important role in the elongation step of protein synthesis.
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC
CC EMBL: X65025; CAA46159.1; --
CC PIR: S22644; R6UTP.
CC InterPro: IPR001813; Ribosomal_60S.
CC InterPro: IPR001859; Ribosomal_P2.
CC Pfam: PF00428; 60S ribosomal; 1.
CC PRINTS: PR00456; RIBOSOMALP2.
CC Ribosomal protein; Phosphorylation.
CC SEQUENCE 109 AA; 10754 MW; D674D31C102F70AF CRC64;
Query Match 1.6%; Score 11; DB 1; Length 109;
Best Local Similarity 100.0%; Pred.No. 0.038;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 507 AAAAAA 517
DB 82 AAAAAA 92
RESULT 2
ID RLAI EIMTE STANDARD; PRT; 114 AA.
AC Q96799;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 60S acidic ribosomal protein P2.
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
OX NCBI_TaxID=5802;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAPE38;
RA Labbe M., Pery P.;
RT "Molecular cloning of a cDNA encoding an acidic ribosomal protein P2
of Eimeria tenella."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Plays an important role in the elongation step of
protein synthesis (By similarity).
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit
(By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF353514; AAK3885.1; ALT INIT.
DR InterPro: IPR001813; Ribosomal 60S.
DR Pfam: PF00428; 60S ribosomal; 1
KW Ribosomal protein; Phosphorylation.
SQ SEQUENCE 114 AA; 11444 MW; 4C083C569078AA9 CRC64;

Query Match 1.6%; Score 11; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAAAATAAAA 517
DB 71 AAAAAAATAAAA 81

RESULT 3
RLAO CAEEL STANDARD; PRT; 311 AA.
AC Q93572;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein P0.
GN RPA-0 OR F25H2.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-15.
RC STRAIN=Bristol N2;
RX MEDLINE=97295299; PubMed=9150941;
RA Bini L., Heid H., Liberatori S., Geier G., Pallini V., Zwillig R.;
RT "Two-dimensional gel electrophoresis of Caenorhabditis elegans
homogenates and identification of protein spots by microsequencing."
RL Electrophoresis 18:557-562(1997).
CC -!- FUNCTION: Ribosomal protein P0 is the functional equivalent of
E.coli protein L10.
CC -!- SUBUNIT: P0 forms a pentameric complex by interaction with dimers
of P1 and P2.
CC -!- SIMILARITY: Belongs to the L10P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z79754; CA502098.1; -
DR PIR; T21351; T21351.
DR Sienna-2DPAGE; Q93572; -
DR WormPep; P25H2.10; CE09655.
DR InterPro: IPR001813; Ribosomal 60S.
DR InterPro: IPR001790; Ribosomal L10.
DR Pfam: PF00428; 60S ribosomal; 1.
DR Pfam: PF00466; Ribosomal L10; 1.
KW Ribosomal protein; Phosphorylation.
XT INIT MET 0
SQ SEQUENCE 311 AA; 33642 MW; 2FA9A35CD24DE0F4 CRC64;

Query Match 1.6%; Score 11; DB 1; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.09;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 AAAAAAATAAAA 518
DB 275 AAAAAAATAAAA 285

RESULT 4
ODP2 ALCEU STANDARD; PRT; 553 AA.
ID ODP2 ALCEU STANDARD; PRT; 553 AA.
AC Q59098;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase
complex (EC 2.3.1.12) [2].
GN PDHB.
OS Alkaligenes eutrophus (Ralstonia eutropha).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H16 / DSM 428 / ATCC 17699;
RX MEDLINE=94292470; PubMed=8021225;
RA Hein S., Steinbuechel A.;
RT "Biochemical and molecular characterization of the Alkaligenes
eutrophus pyruvate dehydrogenase complex and identification of a new
type of dihydrolipoamide dehydrogenase."
RL J. Bacteriol. 175:4394-4408(1994).
CC -!- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall
conversion of pyruvate to acetyl-CoA and CO(2). It contains
multiple copies of three enzymatic components: pyruvate
dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and
lipoamide dehydrogenase (E3).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide = CoA + S-
acetyldihydrolipoamide.
CC -!- COFACTOR: Contains 2 covalently bound lipoyl cofactors (By
similarity).
CC -!- SUBUNIT: Forms a 24-polypeptide structural core with octahedral
symmetry.
CC -!- SIMILARITY: Belongs to the 2-oxoacid dehydrogenase family.
CC -!- SIMILARITY: Contains 2 lipoyl-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC

```

C or send an email to license@isb-sib.ch).
R EXEL; U09865; AAA21599.1; -.
R PIR; B55514; B55514.
R HSP; P10802; IDPC.
R InterPro; IPR001078; 2Oxoacid_dh.
R InterPro; IPR006256; Acef.
R InterPro; IPR000089; Biotin_lipoyl.
R InterPro; IPR003016; Lipoyl_Bs.
R Pfam; PF00198; 2-oxoacid_dh_1.
R Pfam; PF00384; biotin_lipoyl_2.
R ProDom; PD00115; 2Oxoacid_dh; 1.
R TIGRFS; TIGR01348; PDHac_trf_long; 1.
R PROSITE; PS00189; LIPOYL; 2.
W GlycoLysis; Transferase; Acyltransferase; Repeat; Lipoyl.
T BINDING 44 44 LIPOYL (BY SIMILARITY).
T BINDING 162 162 LIPOYL (BY SIMILARITY).
T ACT SITE 526 526 POTENTIAL.
Q SEQUENCE 553 AA; 57338 MW; BF5D370CC60C3F12 CRC64;

Query Match 1.6%; Score 11; DB 1; Length 553;
Best Local Similarity 100.0%; Pred.No. 0.14;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAAAPAAAA 516
b 199 AAAAAPAAAA 209

RESULT 5
D ANPA_PSEAM STANDARD; PRT; 82 AA.
C P04002;
T 23-OCT-1986 (Rel. 02, Created)
T 23-OCT-1986 (Rel. 02, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Antifreeze protein A/B precursor.
S Pseudopleuronectes americanus (winter flounder) (Pleuronectes americanus).
C Eukaryota; Metazoa; Chordata; Vertebrata; Ruteleostomi;
C Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
C Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;
C Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
X NCBI_TaxID=8265;
N [1]
N SEQUENCE FROM N.A. (PROTEIN A).
P MEDLINE=82157490; PubMed=6952188;
A Davies P.L., Roach A.H., Hew C.-L.;
T "DNA sequence coding for an antifreeze protein precursor from winter flounder.";
L Proc. Natl. Acad. Sci. U.S.A. 79:335-339 (1982).
N [2]
P SEQUENCE FROM N.A. (PROTEIN A).
X MEDLINE=88259236; PubMed=3133486;
A Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
T "Differential amplification of antifreeze protein genes in the pleuronectinae.";
L J. Mol. Evol. 27:29-35 (1988).
N [3]
P SEQUENCE FROM N.A. (PROTEIN B).
X MEDLINE=84264559; PubMed=6086629;
A Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
T "Antifreeze protein genes of the winter flounder.";
L J. Biol. Chem. 259:9241-9247 (1984).
N [4]
P SEQUENCE FROM N.A.
X MEDLINE=92209995; PubMed=1555765;
A Davies P.L.;
T "Conservation of antifreeze protein-encoding genes in tandem repeats.";
L Gene 112:163-170 (1992).
N [5]
P 3D-STRUCTURE MODELING OF 45-81.

```

```

RX MEDLINE=92148833; PubMed=1738160;
RA Chou K.-C.;
RT "Energy-Optimized structure of antifreeze protein and its binding mechanism.";
RL J. Mol. Biol. 223:509-517 (1992).
RN [6]
RX X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
RA MEDLINE=95281060; PubMed=7760940;
RA Siccheri F., Yang D.S.C.;
RT "Ice-binding structure and mechanism of an antifreeze protein from winter flounder.";
RL Nature 375:427-431 (1995).
CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point.
CC -!- MISCELLANEOUS: The sequence shown is that of protein A.
CC -!- SIMILARITY: BELONGS TO THE TYPB-I AFP FAMILY. TYPE 1 AFP ARE ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EXBL; L00138; AAB59964.1; -.
DR EXBL; L29178; AAB59964.1; JOINED.
DR EXBL; M62414; AAA49469.1; -.
DR EXBL; X07506; CAA30389.1; -.
DR EXBL; M62416; AAA49471.1; -.
DR EXBL; M62417; AAA49472.1; -.
DR PIR; A05161; A05161.
DR PIR; JS0704; FDFLAW.
DR PIR; JS0706; JS0706.
DR PIR; S02326; S02326.
DR PDB; 1ATF; 15-OCT-94.
DR PDB; 1WFA; 03-JUN-95.
DR PDB; 1WFB; 03-JUN-95.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR0308; ANTIFREEZE1.
DR Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
FT SIGNAL 1 21
FT PROPEP 22 44
FT CHAIN 45 82
FT VARIANT 36 36
FT VARIANT 70 70
FT CONFLICT 24 24
FT HELIX 46 80
SQ SEQUENCE 82 AA; 7711 MW; C2AE7B74C0D46CC1 CRC64;
Query Match 1.5%; Score 10; DB 1; Length 82;
Best Local Similarity 100.0%; Pred.No. 0.22;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 507 AAAAAPAAAA 516
Db 34 AAAAAPAAAA 43

RESULT 6
RL12_METH STANDARD; PRT; 101 AA.
AC P05394;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L12p ('A' type).
GN RPL12P OR MTH1682.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;

```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RA MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safet H., Fatwell D., Prabhakar S.,
RA Mbougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA deltaH: functional analysis and comparative genomics.";
RA J. Bacteriol. 179:7135-7155 (1997).
RN [2]
RP SEQUENCE OF 1-48.
RC MEDLINE=8110590; PubMed=7006702;
RA Matheson A.T., Yaguchi M., Balch W.B., Wolfe R.S.;
RA "Sequence homologies in the N-terminal region of the ribosomal 'A'
RA proteins from Methanobacterium thermoautotrophicum and Halobacterium
RA cutirubrum.";
RA Biochim. Biophys. Acta 626:162-169 (1980).
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
RC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
RC
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RC or send an email to license@isb-sib.ch).
RC
RC Eukaryota; Mycetozoa; Dictyostelium; Dictyostelium.
RC NCBI_TaxID=44689;
RC EMBL: A3000926; AAB86154.1;
RC PIR: H69091; H69091.
RC InterPro: IPR001813; Ribosomal_60S.
RC Pfam: PF00428; 60s_ribosomal; 1.
RC Ribosomal protein; Complete proteome.
RC SEQUENCE 101 AA; 10480 MW; 6C1CEA3BF121FB85 CRC64;
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safet H., Fatwell D., Prabhakar S.,
RA Mbougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA deltaH: functional analysis and comparative genomics.";
RA J. Bacteriol. 179:7135-7155 (1997).
RN [4]
RP SEQUENCE OF 1-48.
RC MEDLINE=8110590; PubMed=7006702;
RA Matheson A.T., Yaguchi M., Balch W.B., Wolfe R.S.;
RA "Sequence homologies in the N-terminal region of the ribosomal 'A'
RA proteins from Methanobacterium thermoautotrophicum and Halobacterium
RA cutirubrum.";
RA Biochim. Biophys. Acta 626:162-169 (1980).
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
RC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
RC
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RC or send an email to license@isb-sib.ch).
RC
RC Eukaryota; Mycetozoa; Dictyostelium; Dictyostelium.
RC NCBI_TaxID=44689;
RC EMBL: A3000926; AAB86154.1;
RC PIR: H69091; H69091.
RC InterPro: IPR001813; Ribosomal_60S.
RC Pfam: PF00428; 60s_ribosomal; 1.
RC Ribosomal protein; Complete proteome.
RC SEQUENCE 101 AA; 10480 MW; 6C1CEA3BF121FB85 CRC64;
RN [5]
RP SEQUENCE FROM N.A.
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safet H., Fatwell D., Prabhakar S.,
RA Mbougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA deltaH: functional analysis and comparative genomics.";
RA J. Bacteriol. 179:7135-7155 (1997).
RN [6]
RP SEQUENCE OF 1-48.
RC MEDLINE=8110590; PubMed=7006702;
RA Matheson A.T., Yaguchi M., Balch W.B., Wolfe R.S.;
RA "Sequence homologies in the N-terminal region of the ribosomal 'A'
RA proteins from Methanobacterium thermoautotrophicum and Halobacterium
RA cutirubrum.";
RA Biochim. Biophys. Acta 626:162-169 (1980).
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
RC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
RC
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RC or send an email to license@isb-sib.ch).
RC
RC Eukaryota; Mycetozoa; Dictyostelium; Dictyostelium.
RC NCBI_TaxID=44689;
RC EMBL: A3000926; AAB86154.1;
RC PIR: H69091; H69091.
RC InterPro: IPR001813; Ribosomal_60S.
RC Pfam: PF00428; 60s_ribosomal; 1.
RC Ribosomal protein; Complete proteome.
RC SEQUENCE 101 AA; 10480 MW; 6C1CEA3BF121FB85 CRC64;
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF044919; AAC64512.1;
CC InterPro: IPR001813; Ribosomal_60S.
CC Pfam: PF00428; 60s_ribosomal; 1.
CC Ribosomal protein.
CC SEQUENCE 101 AA; 10292 MW; 27262815729C8CB0 CRC64;
RN [7]
RP SEQUENCE FROM N.A.
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safet H., Fatwell D., Prabhakar S.,
RA Mbougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA deltaH: functional analysis and comparative genomics.";
RA J. Bacteriol. 179:7135-7155 (1997).
RN [8]
RP SEQUENCE OF 1-48.
RC MEDLINE=8110590; PubMed=7006702;
RA Matheson A.T., Yaguchi M., Balch W.B., Wolfe R.S.;
RA "Sequence homologies in the N-terminal region of the ribosomal 'A'
RA proteins from Methanobacterium thermoautotrophicum and Halobacterium
RA cutirubrum.";
RA Biochim. Biophys. Acta 626:162-169 (1980).
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
RC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
RC
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RC or send an email to license@isb-sib.ch).
RC
RC Eukaryota; Mycetozoa; Dictyostelium; Dictyostelium.
RC NCBI_TaxID=44689;
RC EMBL: A3000926; AAB86154.1;
RC PIR: H69091; H69091.
RC InterPro: IPR001813; Ribosomal_60S.
RC Pfam: PF00428; 60s_ribosomal; 1.
RC Ribosomal protein; Complete proteome.
RC SEQUENCE 101 AA; 10480 MW; 6C1CEA3BF121FB85 CRC64;
RN [9]
RP SEQUENCE FROM N.A.
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safet H., Fatwell D., Prabhakar S.,
RA Mbougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA deltaH: functional analysis and comparative genomics.";
RA J. Bacteriol. 179:7135-7155 (1997).
RN [10]
RP SEQUENCE OF 1-48.
RC MEDLINE=8110590; PubMed=7006702;
RA Matheson A.T., Yaguchi M., Balch W.B., Wolfe R.S.;
RA "Sequence homologies in the N-terminal region of the ribosomal 'A'
RA proteins from Methanobacterium thermoautotrophicum and Halobacterium
RA cutirubrum.";
RA Biochim. Biophys. Acta 626:162-169 (1980).
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
RC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
RC
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RC or send an email to license@isb-sib.ch).
RC
RC Eukaryota; Mycetozoa; Dictyostelium; Dictyostelium.
RC NCBI_TaxID=44689;
RC EMBL: A3000926; AAB86154.1;
RC PIR: H69091; H69091.
RC InterPro: IPR001813; Ribosomal_60S.
RC Pfam: PF00428; 60s_ribosomal; 1.
RC Ribosomal protein; Complete proteome.
RC SEQUENCE 101 AA; 10480 MW; 6C1CEA3BF121FB85 CRC64;
```

```
Query Match 1.5%; Score 10; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. NO. 0.26; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 506 VAAAAAPAAA 515
DB 61 VAAAAAPAAA 70

RESULT 8
RL12 METTL STANDARD; PRT; 105 AA.
AC P22683;
AT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 60S acidic ribosomal protein P2.
GN RPLP2.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelium; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=AX3;
RC MEDLINE=91232921; PubMed=1840653;
RA Prieto J., Candel E., Coloma A.;
RT "Nucleotide sequence of a cDNA encoding acidic ribosomal
RL phosphoprotein P2 in Dictyostelium discoideum.";
RL Nucleic Acids Res. 19:1341-1341(1991).
RC -!- FUNCTION: Plays an important role in the elongation step of
CC protein synthesis.
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X56192; CAA39655.1;
CC PIR: S14014; R6DOP2.
CC DictyBase: DB0001822; X56192.
CC InterPro: IPR001813; Ribosomal_60S.
CC Pfam: PF00428; 60s_ribosomal; 1.
CC Ribosomal protein; Phosphorylation.
CC INIT MET 0
CC SEQUENCE 105 AA; 10378 MW; 519FAB6679A5B840 CRC64;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safet H., Fatwell D., Prabhakar S.,
RA Mbougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA deltaH: functional analysis and comparative genomics.";
RA J. Bacteriol. 179:7135-7155 (1997).
RN [2]
RP SEQUENCE OF 1-48.
RC MEDLINE=8110590; PubMed=7006702;
RA Matheson A.T., Yaguchi M., Balch W.B., Wolfe R.S.;
RA "Sequence homologies in the N-terminal region of the ribosomal 'A'
RA proteins from Methanobacterium thermoautotrophicum and Halobacterium
RA cutirubrum.";
RA Biochim. Biophys. Acta 626:162-169 (1980).
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
RC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
RC
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RC or send an email to license@isb-sib.ch).
RC
RC Eukaryota; Mycetozoa; Dictyostelium; Dictyostelium.
RC NCBI_TaxID=44689;
RC EMBL: A3000926; AAB86154.1;
RC PIR: H69091; H69091.
RC InterPro: IPR001813; Ribosomal_60S.
RC Pfam: PF00428; 60s_ribosomal; 1.
RC Ribosomal protein; Complete proteome.
RC SEQUENCE 101 AA; 10480 MW; 6C1CEA3BF121FB85 CRC64;
```

```
Query Match 1.5%; Score 10; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. NO. 0.27; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 506 VAAAAAPAAA 515
DB 62 VAAAAAPAAA 71

RESULT 8
RL12 METTL STANDARD; PRT; 101 AA.
AC O52706;
AT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2004 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L12P.
GN RPL12P.
OS Methanococcus thermolithotrophicus.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanothermococcus.
OX NCBI_TaxID=2186;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safet H., Fatwell D., Prabhakar S.,
RA Mbougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA deltaH: functional analysis and comparative genomics.";
RA J. Bacteriol. 179:7135-7155 (1997).
RN [2]
RP SEQUENCE OF 1-48.
RC MEDLINE=8110590; PubMed=7006702;
RA Matheson A.T., Yaguchi M., Balch W.B., Wolfe R.S.;
RA "Sequence homologies in the N-terminal region of the ribosomal 'A'
RA proteins from Methanobacterium thermoautotrophicum and Halobacterium
RA cutirubrum.";
RA Biochim. Biophys. Acta 626:162-169 (1980).
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
RC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
RC
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RC or send an email to license@isb-sib.ch).
RC
RC Eukaryota; Mycetozoa; Dictyostelium; Dictyostelium.
RC NCBI_TaxID=44689;
RC EMBL: A3000926; AAB86154.1;
RC PIR: H69091; H69091.
RC InterPro: IPR001813; Ribosomal_60S.
RC Pfam: PF00428; 60s_ribosomal; 1.
RC Ribosomal protein; Complete proteome.
RC SEQUENCE 101 AA; 10480 MW; 6C1CEA3BF121FB85 CRC64;
```



```

GN RPLL OR LMO0251.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN (1)_SEQUENCE FROM N.A.
RP STRAIN=EGD-e / Serovar 1/2a;
RC MEDLINE=21537279; PubMed=11679669;
RX Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero P., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann B., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -!- FUNCTION: Seems to be the binding site for several of the factors
CC involved in protein synthesis and appears to be essential for
CC accurate translation (By similarity).
CC -!- SIMILARITY: Belongs to the L12p family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AL591974; CAD00778.1; -
CC PIR; AD1106; AD1106.
CC ListList; LMO0251; -
CC HAMAP; MF_00368; -; 1.
CC InterPro; IPR008932; Ribos_L12/7_olig.
CC InterPro; IPR000206; Ribosomal_L12.
CC Pfam; PF00542; Ribosomal_L12; 1.
CC ProDom; PD001326; Ribosomal_L12; 1.
CC TIGRFAMs; TIGR00855; L12; 1.
CC Ribosomal protein; Complete proteome.
CC KW SEQUENCE 120 AA; 12469 MW; 8EC3D5417F6E1954 CRC64;
CC
CC Query Match 1.5%; Score 10; DB 1; Length 120;
CC Best Local Similarity 100.0%; Pred.No. 0.3;
CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 500 AAAPVAVAAA 509
CC Db |||||
CC 35 AAAPVAVAAA 44
CC
CC RESULT 11
CC RL7_STRPN STANDARD; PRT; 121 AA.
CC ID RL7_STRPN
CC AC P80714;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE 50S ribosomal protein L7/L12.
CC GN RPLL OR SPL354 OR SPL311.
CC OS Streptococcus pneumoniae, and
CC OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CC OX NCBI_TaxID=1313, 171101;
CC (1)
CC RN SEQUENCE FROM N.A.
CC RP STRAIN=ATCC BAA-334 / TIGR4;
CC MEDLINE=21357209; PubMed=11463916;

```

```

RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Whitt M., Kolonay J.P., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salanay S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey B.K.,
RA Holt I.E., Loftus B.J., Yang P., Smith H.O., Venter J.C., Hickey B.K.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Aborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAburn S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peary R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
RN [3]
RP SEQUENCE OF 1-42.
RX MEDLINE=9717772; PubMed=9025278;
RA Kolberg J., Heiby E.A., Lopez R., Sletten K.;
RT "Monoclonal antibodies against Streptococcus pneumoniae detect
RT epitopes on eubacterial ribosomal proteins L7/L12 and on
RT streptococcal elongation factor Ts.";
RL Microbiology 143:55-61(1997).
CC -!- FUNCTION: Seems to be the binding site for several of the factors
CC involved in protein synthesis and appears to be essential for
CC accurate translation (By similarity);
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE010507; AAL94122.1; -.
CC HAMAP: MF_00368; -. 1.
CC InterPro: IPR008932; Ribos_L12/7_olig.
CC DR Pfam: IPR000206; Ribosomal_L12.
CC DR Pfam: PF00542; Ribosomal_L12; 1.
CC DR ProDom: PD001326; Ribosomal_L12; 1.
CC DR TIGRFAMs: TIGR00855; L12; 1.
CC KW Ribosomal protein; Complete proteome.
CC INIT_MET 0
SQ SEQUENCE 121 AA; 12605 MW; 4382328A2B266BA3 CRC64;

Query Match 1.5%; Score 10; DB 1; Length 121;
Best Local Similarity 100.0%; Pred.No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 13
RL7_BRUME STANDARD; PRT; 124 AA.
AC P41106; Q9R2F1;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPLL OR RME10748 OR BR1245.
OS Brucella melitensis,
OS Brucella suis, and
OS Brucella abortus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459, 29461, 235;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B.melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=94357413; PubMed=8076798;

```

```

AC Q8RH15;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RELL OR FN2037.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11893109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Fusch G., Haselkorn R.,
RA Fongsteine M., Kyridides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- FUNCTION: Seems to be the binding site for several of the factors
CC involved in protein synthesis and appears to be essential for
CC accurate translation (By similarity);
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE010507; AAL94122.1; -.
CC HAMAP: MF_00368; -. 1.
CC InterPro: IPR008932; Ribos_L12/7_olig.
CC DR Pfam: IPR000206; Ribosomal_L12.
CC DR Pfam: PF00542; Ribosomal_L12; 1.
CC DR ProDom: PD001326; Ribosomal_L12; 1.
CC DR TIGRFAMs: TIGR00855; L12; 1.
CC KW Ribosomal protein; Complete proteome.
CC INIT_MET 0
SQ SEQUENCE 122 AA; 12605 MW; 4382328A2B266BA3 CRC64;

Query Match 1.5%; Score 10; DB 1; Length 122;
Best Local Similarity 100.0%; Pred.No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 13
RL7_BRUME STANDARD; PRT; 124 AA.
AC P41106; Q9R2F1;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPLL OR RME10748 OR BR1245.
OS Brucella melitensis,
OS Brucella suis, and
OS Brucella abortus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459, 29461, 235;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B.melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=94357413; PubMed=8076798;

```


Search completed: June 16, 2004, 19:26:11
Job time : 18 secs

```
DR HAMAP: MF_00368; -: 1.
DR InterPro: IPR008932; Ribos_L12/7 olig.
DR InterPro: IPR000206; Ribosomal_L12.
DR Pfam: PF00542; Ribosomal_L12; 1.
DR ProDom: PD001326; Ribosomal_L12; 1.
DR TIGRFAMs: TIGR00855; L12; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 125 AA; 12736 MW; 37A8E02FE77175D CRC64;

Query Match 1.5%; Score 10; DB 1; Length 125;
Best Local Similarity 100.0%; Pred. NO. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 500 AAAPVAVAAA 509
2b 34 AAAPVAVAAA 43

RESULT 15
2L7_RHILQ
ID - RL7_RHILQ STANDARD; PRT; 125 AA.
AC 098N67;
YT 28-FEB-2003 (Rel. 41, Created)
YT 28-FEB-2003 (Rel. 41, Last sequence update)
YT 28-FEB-2003 (Rel. 41, Last annotation update)
2E 50S ribosomal protein L7/L12.
2N RPLL OR MLR0275.
2S Rhizobium loti (Mesorhizobium loti).
2C Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
2C Phyllobacteriaceae; Mesorhizobium.
2X NCBI_TaxID=381;
2N (1)
SEQUENCE FROM N.A.
2X STRAIN=MAFF303099;
2X MEDLINE=21082930; PubMed=11214968;
2A Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
2A Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
2A Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
2A Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
2A Takeuchi C., Yamada M., Tabata S.;
2A "Complete genome structure of the nitrogen-fixing symbiotic bacterium
2T Mesorhizobium loti."
2L DNA Res. 7:331-338(2000).
2C -!- FUNCTION: Seems to be the binding site for several of the factors
2C involved in protein synthesis and appears to be essential for
2C accurate translation (by similarity).
2C -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
2C -----
2C This SWISS-PROT entry is copyright. It is produced through a collaboration
2C between the Swiss Institute of Bioinformatics and the EMBL outstation -
2C the European Bioinformatics Institute. There are no restrictions on its
2C use by non-profit institutions as long as its content is in no way
2C modified and this statement is not removed. Usage by and for commercial
2C entities requires a license agreement (see http://www.isb-sib.ch/announce/
2C or send an email to license@isb-sib.ch).
2C -----
2C EMBL: AP002994; BAB47895.1; -.
2R HAMAP: MF_00368; -: 1.
2R InterPro: IPR008932; Ribos_L12/7 olig.
2R InterPro: IPR000206; Ribosomal_L12.
2R Pfam: PF00542; Ribosomal_L12; 1.
2R ProDom: PD001326; Ribosomal_L12; 1.
2R TIGRFAMs: TIGR00855; L12; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 125 AA; 12735 MW; 9EE14454212A55A5P CRC64;

Query Match 1.5%; Score 10; DB 1; Length 125;
Best Local Similarity 100.0%; Pred. NO. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 500 AAAPVAVAAA 509
2b 34 AAAPVAVAAA 43
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: June 16, 2004, 19:15:43 ; Search time 61 Seconds
(without alignments)
3126.550 Million cell updates/sec

Title: US-09-332-063-2
Perfect score: 675
Sequence: 1 MPRAQSSASYQVPADPPA.....KTPILQIGSPDAENVYLI 675

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	675	100.0	675	3	AAY54052 An angiotensin
2	525	77.8	675	3	AAY54053 A variant
3	492	72.9	993	7	AAB37918 Human CDP
4	479	71.0	479	5	ABP43965 Unidentif
5	183	27.1	759	4	ABG17959 Novel hum
6	143	21.2	143	3	AAY54054 Angiotat
7	55	8.1	208	4	AAM16746 Peptide #
8	55	8.1	208	4	ABB35732 Peptide #
9	55	8.1	208	4	AAM29234 Peptide #
10	55	8.1	208	4	ABB30567 Peptide #
11	55	8.1	208	4	ABB21158 Protein #
12	55	8.1	208	4	AAM68923 Human bon
13	55	8.1	208	4	AAM56541 Human bra
14	55	8.1	208	4	ABG50588 Human liv
15	55	8.1	208	4	AAM04463 Peptide #
16	55	8.1	208	5	ABG38504 Human pep
17	48	7.1	48	4	AAM19612 Peptide #
18	48	7.1	48	4	ABB39256 Peptide #
19	48	7.1	48	4	AAM32758 Peptide #
20	48	7.1	48	4	ABB24091 Protein #
21	48	7.1	48	4	AAM72514 Human bon
22	48	7.1	48	4	AAM59920 Human bra
23	48	7.1	48	4	ABG54205 Human liv
24	48	7.1	48	5	ABG42334 Human pep
25	26	3.9	26	4	AAM33005 Peptide #

26	26	3.9	26	4	AAM72776 Human bon
27	26	3.9	26	4	AAM60161 Human bra
28	26	3.9	26	4	ABG54477 Human liv
29	26	3.9	26	5	ABG42601 Human pep
30	24	3.6	608	4	AAU69439 Human pur
31	24	3.6	637	5	ABP69085 Human pol
32	24	3.6	801	7	ADC38518 Human AML
33	24	3.6	869	7	ADC37654 Human ang
34	24	3.6	882	7	ABR56262 Mouse Jun
35	18	2.7	340	4	ABU70751 Human adi
36	18	2.7	340	4	AAM93770 Human pol
37	18	2.7	467	5	ABG97502 Human NOV
38	18	2.7	772	7	ABR56263 Mouse Jun
39	18	2.7	849	3	AAB41861 Human ORP
40	13	1.9	562	7	ADC37657 Human ang
41	13	1.9	706	4	ABR32866 Human pro
42	11	1.6	16	2	AAM32225 A polyala
43	11	1.6	76	2	AAR22394 Antigen t
44	11	1.6	228	6	ABO00600 Novel hum
45	10	1.5	82	2	AAY23877 Winter fl

ALIGNMENTS

RESULT 1

AAY54052

ID AAY54052 standard; protein; 675 AA.

XX AC AAY54052;

XX DT 27-MAR-2000 (first entry)

XX DE An angiogenesis-associated protein which binds plasminogen.

XX KW Human; angiogenesis-associated protein; plasminogen; ABP-1;

XX KW kringle domain; angiotensin; plasminogen receptor;

XX KW angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;

XX KW inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;

XX KW obesity; Gene therapy.

XX OS Homo sapiens.

XX PN WO9966038-A1.

XX PD 23-DEC-1999.

XX PF 11-JUN-1999; 99WO-EP004109.

XX PR 15-JUN-1998; 98SE-00002130.

XX PR 15-JUN-1998; 98US-0089466P.

XX PR 17-DEC-1998; 98SE-00004372.

XX PR 29-DEC-1998; 98US-0114386P.

XX PR (PAA) PHARMACIA & UPJOHN AB.

XX PI Holmgren L, Troyanovsky B;

XX DR WPI; 2000-106099/09.

XX DR N-PSDB; AAZ45329.

XX PT Novel human protein useful for treating angiogenesis associated diseases

XX PT or disorders.

XX PS Claim 4; Page 43-46; 58pp; English.

XX PS The present sequence represents a human angiogenesis-associated protein

XX CC which is able to bind an N-terminal fragment of plasminogen. The protein

XX CC is designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or

XX CC kringle5 (K5) of plasminogen. These four kringle domains comprise

XX CC angiotensin. The protein acts as a receptor for plasminogen. The

XX CC angiotensin-binding domain of the ABP-1 protein is described in AAY54054.

XX CC A polymorphic variant of ABP-1 is also described, in AAY54053. ABP-1 can

CC be used to manufacture medicaments for treating angiogenesis-related
CC diseases or disorders, such as tumor conditions, diabetes, rheumatoid
CC arthritis, and even some inflammatory diseases such as psoriasis. Chronic
CC inflammation of the intestine, asthma, etc. . The protein may also be able
CC to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene
CC therapy techniques
CX
SQ Sequence 675 AA;

Query Match 100.0%; Score 675; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 MPRAQSSASVOPADPFAIVSAQOMVILSDENNLFOELGCGVKEVARLQKVETRI 60
2b 1 MPRAQSSASVOPADPFAIVSAQOMVILSDENNLFOELGCGVKEVARLQKVETRI 60

2Y 61 QRVSEAVENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 120
2b 61 QRVSEAVENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 120

2Y 121 SEDTRKTIISQFAKNKESOREKEKLEAELATARNSTNEDQRRHLEIRDOALSNAQAKVVKL 180
2b 121 SEDTRKTIISQFAKNKESOREKEKLEAELATARNSTNEDQRRHLEIRDOALSNAQAKVVKL 180

2Y 181 EELKKQVYVDKVKVXQQAALVOLQAACEKEQLEHLRLTRLERELSLRIQORQNCQP 240
2b 181 EELKKQVYVDKVKVXQQAALVOLQAACEKEQLEHLRLTRLERELSLRIQORQNCQP 240

2Y 241 TNVSEYNAAALMELLREKEERILALEADMTKWEQKYLEENVMHFALDAATAVAARDIT 300
2b 241 TNVSEYNAAALMELLREKEERILALEADMTKWEQKYLEENVMHFALDAATAVAARDIT 300

2Y 301 VISHSPNTSYDTALEARIOKEEERILMANKECLDMGRIKTLHAQIIEKDAMIKVLQORS 360
2b 301 VISHSPNTSYDTALEARIOKEEERILMANKECLDMGRIKTLHAQIIEKDAMIKVLQORS 360

2Y 361 RKEPSKTQLSCMPAPKSLMSISNAGSGLLSHSSTLTGSPIMEKRRDDKSWKSGSLIGL 420
2b 361 RKEPSKTQLSCMPAPKSLMSISNAGSGLLSHSSTLTGSPIMEKRRDDKSWKSGSLIGL 420

2Y 421 GDYRAEYVSPSPVPPSTPLLSAHSKTKGSDCSTQTERGTESNKTAAVAPISVPAPVAA 480
2b 421 GDYRAEYVSPSPVPPSTPLLSAHSKTKGSDCSTQTERGTESNKTAAVAPISVPAPVAA 480

2Y 481 AATAAALTATAATITTTWVAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAAGQIPA 540
2b 481 AATAAALTATAATITTTWVAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAAGQIPA 540

2Y 541 AASVASAAAVAPSAASAAAVAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 600
2b 541 AASVASAAAVAPSAASAAAVAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 600

2Y 601 VAPTPTPTTFAVAQAEVSPASPTGPGPHRLSIPSLTCNPKDGTGPFVHNTLERTPIQ 660
2b 601 VAPTPTPTTFAVAQAEVSPASPTGPGPHRLSIPSLTCNPKDGTGPFVHNTLERTPIQ 660

2Y 661 ILGQEPDAEWVEYLI 675
2b 661 ILGQEPDAEWVEYLI 675

RESULT 2
ID AAY54053
XX AAY54053 standard; protein; 675 AA.
AC AAY54053;
XX
XX
XX 27-MAR-2000 (first entry)
DE A variant of an angiogenesis-associated protein which binds plasminogen.
XX Human; angiogenesis-associated protein; plasminogen; ABP-1;

KW kringle domain; angiotatin; plasminogen receptor;
KW angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
KW inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
XX obesity; gene therapy.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT Misc-difference 135
FT Misc-difference /label= Asn, Ser, Asp
FT Misc-difference 148..150
FT /note= "these residues are either Glu-Leu-Ala or Thr-Thr-Pro"
XX WO9966038-A1.
XX 23-DEC-1999.
XX 11-JUN-1999; 99WO-BP004109.
XX 15-JUN-1998; 98SE-00002130.
XX 15-JUN-1998; 98US-0089266P.
XX 17-DEC-1998; 98SE-00004372.
XX 29-DEC-1998; 98US-0114386P.
XX (PHAA) PHARMACIA & UPJOHN AB.
XX Holmgren L, Troyanovsky B;
XX WPI; 2000-106099/09.
XX Novel human protein useful for treating angiogenesis associated diseases or disorders.
XX Claim 5; Page 46-49; 58pp; English.
XX The present sequence represents a polymorphic variant of a human
XX angiogenesis-associated protein which is able to bind an N-terminal
XX fragment of plasminogen. The native protein, described in AAY54052, is
XX designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or
XX kringle 5 (K5) of plasminogen. These four kringle domains comprise
XX angiotatin. The ABP-1 protein acts as a receptor for plasminogen. The
XX angiotatin-binding domain of the ABP-1 protein is described in AAY54054.
XX ABP-1 can be used to manufacture medicaments for treating angiogenesis-
XX related diseases or disorders, such as tumor conditions, diabetes,
XX rheumatoid arthritis, and even some inflammatory diseases such as
XX psoriasis, chronic inflammation of the intestine, asthma, etc. . The ABP-
XX protein may also be able to treat and cure, or prevent, obesity. The ABP-
XX 1 DNA can be used in gene therapy techniques

Sequence 675 AA;
Query Match 77.8%; Score 525; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 TARSTNEQRRHIEIRDOALSNAQAKVVKLEELKKQVYVDKVKVXQQAALVOLQAACEK 210
DB 151 TARSTNEQRRHIEIRDOALSNAQAKVVKLEELKKQVYVDKVKVXQQAALVOLQAACEK 210

QY 211 REQLEHLRLTRLERELSLRIQORQNCQPTNVSEYNAAALMELLREKEERILALEADMT 270
DB 211 REQLEHLRLTRLERELSLRIQORQNCQPTNVSEYNAAALMELLREKEERILALEADMT 270

QY 271 KWEQKYLEENVMHFALDAATAVAARDITTVISHSPNTSYDTALEARIOKEEERILMAN 330
DB 271 KWEQKYLEENVMHFALDAATAVAARDITTVISHSPNTSYDTALEARIOKEEERILMAN 330

QY 331 RCLDMGRIKTLHAQIIEKDAMIKVLQORSRKESKTEQLSCMRPAKSLMSISNAGSGLL 390
DB 331 RCLDMGRIKTLHAQIIEKDAMIKVLQORSRKESKTEQLSCMRPAKSLMSISNAGSGLL 390

QY 391 SHSSTLTGSPIMEKRRDDKSWKSGSLIGLGDYAEVVPSTPSPVPPSTPLLSAHSKTKGS 450

391 SHSSTLGTSPIMEKXDDKSWKSLGLLLGGDYRAEYVPSPTSPVPSPTLLSAHSKTS 450
 451 RDGSGTQTERGTSTKTAAPAPISVPAPVAAAATAAATTAATITTTTMAAPVAAVAAA 510
 451 RDGSGTQTERGTSTKTAAPAPISVPAPVAAAATAAATTAATITTTTMAAPVAAVAAA 510
 511 AAAAAAPSPATAAATAAASPAAGQIPAAASVASAAVAPASAAAAAVQVAPAPV 570
 511 AAAAAAPSPATAAATAAASPAAGQIPAAASVASAAVAPASAAAAAVQVAPAPV 570
 571 PAPALVVPAPAPAAQASAPATQAPTAPAPVAPAPAPAPAPAPAPAPAPAPAPV 630
 571 PAPALVVPAPAPAAQASAPATQAPTAPAPVAPAPAPAPAPAPAPAPAPAPV 630
 631 LSIPSLTCNPKDGTGVPVHSHNTLERTPIQILGOBPAEAWVEYLI 675
 631 LSIPSLTCNPKDGTGVPVHSHNTLERTPIQILGOBPAEAWVEYLI 675

RESULT 3
 LAE37918
 D AAE37918 standard; protein; 993 AA.
 X X
 C AAE37918;
 T 06-NOV-2003 (first entry)
 E Human CGDD-7 protein.
 X Human; cell growth, differentiation and death protein; CGDD; leukaemia;
 W neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
 W muscular disorder; myotonic dystrophy; catatonia; endocrine disorder;
 W diabetes; Grave's disease; cancer; immunological disorder; scleroderma;
 W systemic lupus erythematosus; allergy; Crohn's disease; renal disorder;
 W gastrointestinal disorder; Goodpasture's syndrome; infection; cirrhosis;
 W cardiovascular disorder; atherosclerosis; hepatic disease; transgenic;
 W transgenic animal; gene therapy; neuroprotective; relaxant; cytostatic;
 W dermatological; immunosuppressive; cerebroprotective; anticonvulsant;
 W antibacterial; antiparasitic; fungicide; virucide; uropathic; cardiant;
 W protozoacide; nootropic.
 X Homo sapiens.
 X X
 X W02003050253-A2.
 X 19-JUN-2003.
 X 04-DEC-2002; 2002WO-US039133.
 X 07-DEC-2001; 2001US-0340747P.
 X 20-DEC-2001; 2001US-0342761P.
 X 15-JAN-2002; 2002US-0349705P.
 X 06-FEB-2002; 2002US-0354764P.
 X 12-FEB-2002; 2002US-0356216P.
 X (INCY-) INCYTE GENOMICS INC.
 X Griffin JA, Runkumar J, Emerling BM, Kable AE, Elliott VS;
 X Marquis JP, Baughn MR, Corvad AB, Yue H, Lee EA, Becha SD, Tang YT;
 X Tran UK, Swarnakar A, Lee S, Ison CH, Hafalia AJA, Tran B;
 X Sprague WW, Lee SY, Khare R, Gandhi AR, Gietzen KJ, Bhatia U;
 X Burrill JD, Blake JJ, Ho A, Zheng W;
 X WPI; 2003-532903/50.
 X N-PSDB; AAD57227.
 X X
 X New CGDD polypeptides, useful for diagnosing, preventing, and treating
 X disorders associated with an abnormal expression or activity of CGDD,
 X e.g. neuromuscular, immunological, cardiovascular disorders, cancer
 X and/or infections.
 X Claim 1; Page 220-222; 299pp; English.

XX The present invention relates to novel cell growth, differentiation and
 CC death (CGDD) proteins and polynucleotides encoding them. The sequences of
 CC the invention are useful in diagnosing, preventing and treating disorders
 CC associated with an abnormal expression or activity of CGDD such as
 CC neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's
 CC disease), muscular disorders (e.g. myotonic dystrophy, catatonia),
 CC endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g.
 CC leukaemia, cervical or breast cancers), immunological disorders (e.g.
 CC scleroderma, systemic lupus erythematosus, allergies), gastrointestinal
 CC disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's
 CC syndrome), infections (e.g. viral, bacterial, fungal, parasitic,
 CC protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis)
 CC and hepatic diseases (e.g. cirrhosis). The polynucleotides can be used to
 CC create humanised animals or transgenic animals to model human diseases.
 CC The invention is also used in gene therapy. The present sequence is human
 CC CGDD-7 protein
 XX Sequence 993 AA;
 SQ
 Query Match 72.9%; Score 492; DB 7; Length 993;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPRAQSSASQYVPADPFAIVSRAQOMVEILSDENRNLROLEGCEYKVARLOKVEI 60
 Db 319 MPRAQSSASQYVPADPFAIVSRAQOMVEILSDENRNLROLEGCEYKVARLOKVEI 378
 QY 61 QRYSEAYENLVKSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEYEG 120
 Db 379 QRYSEAYENLVKSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEYEG 438
 QY 121 SEDTRKTIISQIFANKESQREKELAEIATARNSTNEDQRRHIEIRDOALSNAQKVKL 180
 Db 439 SEDTRKTIISQIFANKESQREKELAEIATARNSTNEDQRRHIEIRDOALSNAQKVKL 498
 QY 181 EERUKKQVTVKVMQOALVOLQACEKKEQLERLRLERLESLRIQORQNCQP 240
 Db 499 EELUKKQVTVKVMQOALVOLQACEKKEQLERLRLERLESLRIQORQNCQP 558
 QY 241 TNVSEYNAALMBELLREKERILALEADMTKWQKYLEBNVMHFPALDAAATVAAQRDTT 300
 Db 559 TNVSEYNAALMBELLREKERILALEADMTKWQKYLEBNVMHFPALDAAATVAAQRDTT 618
 QY 301 VISHSPNTSYDTALEARIQKEEIEILWANKRCLDMEGRIKTLHAQIIIEKDAMIKVQQRS 360
 Db 619 VISHSPNTSYDTALEARIQKEEIEILWANKRCLDMEGRIKTLHAQIIIEKDAMIKVQQRS 678
 QY 361 RKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIMEBKDDKSKWSGLILG 420
 Db 679 RKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIMEBKDDKSKWSGLILG 738
 QY 421 GDYRAEYVPTSPVPPTPELLSAHKTGSRDCSTOTERTSNTKTAANAPISVPAPVAA 480
 Db 739 GDYRAEYVPTSPVPPTPELLSAHKTGSRDCSTOTERTSNTKTAANAPISVPAPVAA 798
 QY 481 AATAAAITATAA 492
 Db 799 AATAAAITATAA 810
 RESULT 4
 ABP43965
 ID ABP43965 standard; protein; 479 AA.
 X X
 AC ABP43965;
 X X
 X 26-FEB-2003 (first entry)
 X Unidentified protein sequence.
 X Neuroprotective; immunomodulator; cancer; chromosome X; cytostatic;
 X anti-inflammatory; gene therapy; nutritional supplement; wound; burn;
 X

KW ulcer; Alzheimer's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW vulnery.
OS Homo sapiens.
PN WO200231111-A2.
PD 18-APR-2002.
XX 11-OCT-2001; 2001WO-US027760.
XX 12-OCT-2000; 2000US-00687527.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX 2001-426278/45.
XX N-PSDB; ABQ61209.
XX New polypeptides and their encoded proteins, useful as nutritional
XX sources or supplements, or in gene therapy, particularly for treating
XX wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
XX inflammation.
XX Claim 20; SEQ ID # 868; 357pp + Sequence Listing; English.
XX The invention relates to 446 newly isolated polynucleotide sequences. The
XX activity of polynucleotides of the invention may be described as,
XX vulnery, neuroprotective, immunomodulator, cytostatic and anti-
XX inflammatory. Compositions comprising nucleic acids of the invention are
XX useful for treating a mammalian subject, or as nutritional sources or
XX supplements. These are useful in gene therapy, particularly for treating
XX wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
XX amyotrophic lateral sclerosis, autoimmune disorders, cancer or
XX inflammation. The nucleic acids and polypeptides are also useful in
XX diagnostic and research methods. The sequences given in records ABP43544-
XX ABP43989 represent polypeptides encoded by polynucleotides of the
XX invention. NOTE: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 479 AA;

Query Match
Best Local Similarity 71.0%; Score 479; DB 5; Length 479;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 MQQALVQLQAACEKREQLERLRLERLESLRIQQRGNCQPTNVSEYNAALMELLR 256
DB 1 MQQALVQLQAACEKREQLERLRLERLESLRIQQRGNCQPTNVSEYNAALMELLR 60
QY 257 EKEERILALEADMTKWEQKYLEENVMRHFALDAAATVAARDTTVISHSPNTSDTALEA 316
DB 61 EKEERILALEADMTKWEQKYLEENVMRHFALDAAATVAARDTTVISHSPNTSDTALEA 120
QY 317 RIQKEEERILANKCELDMEGRITKHAQIEKDAIKVLQORSRKEPSTQLSCMRPA 376
DB 121 RIQKEEERILANKCELDMEGRITKHAQIEKDAIKVLQORSRKEPSTQLSCMRPA 180
QY 377 KSLMISNAGSGLLSHSSTLTGSPIMBEKRDKSKWKSIGLILGDDVRAEYVSTSPVP 436
DB 181 KSLMISNAGSGLLSHSSTLTGSPIMBEKRDKSKWKSIGLILGDDVRAEYVSTSPVP 240
QY 437 PSTPLLSAHSKSGRDCSTQTERGTSNKTAAPVAPVAPVAAAATAATATAATITT 496
DB 241 PSTPLLSAHSKSGRDCSTQTERGTSNKTAAPVAPVAPVAAAATAATATAATITT 300
QY 497 TWVAAAPVAVAAAAAPAAAAAPSPATAATAATAATAATAATAATAATAATAATA 556
DB 301 TWVAAAPVAVAAAAAPAAAAAPSPATAATAATAATAATAATAATAATAATAATA 360

QY 557 AAAVQVAPAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 616
DB 361 AAAVQVAPAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 420
QY 617 EVRASPTGPGPHRLSPSLTCNPKDGTGPFVFSNLTERTPIQILQGBDAEWVEYLI 675
DB 421 EVRASPTGPGPHRLSPSLTCNPKDGTGPFVFSNLTERTPIQILQGBDAEWVEYLI 479

RESULT 5
ABG17959
ID ABG17959 standard; protein; 759 AA.
XX ABG17959;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #17950.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS82146.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 20; SEQ ID NO 48318; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 759 AA;

Query Match
27.1%; Score 183; DB 4; Length 759;

Best Local Similarity 100.0%; Pred. No. 8.5e-157;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPRAQPSASQVPADPFAIVSRAQWVEILSDENRNLQEGCGYKVARLQKVETEI 60
DB 525 MPRAQPSASQVPADPFAIVSRAQWVEILSDENRNLQEGCGYKVARLQKVETEI 584
QY 61 QRVSEAYENLVKSSKRALEKAVNKLGEIRWHDFFNDRDLERLETANKOLAEKEYEG 120
DB 585 QRVSEAYENLVKSSKRALEKAVNKLGEIRWHDFFNDRDLERLETANKOLAEKEYEG 644
QY 121 SEDTRKTIQSILFAKNKESQREKELEAEELATARGTNEQRRHIEIRQALSNAQAKVVKL 180
DB 645 SEDTRKTIQSILFAKNKESQREKELEAEELATARGTNEQRRHIEIRQALSNAQAKVVKL 704
QY 181 EEE 183
DB 705 EEE 707

RESULT 6
AAV54054
ID AAV54054 standard; protein; 143 AA.
AC AAV54054;
DT 27-MAR-2000 (first entry)
CX Angiostatin-binding domain of ABP-1, designated Big-3.
KW Human; angiogenesis-associated protein; plasminogen; ABP-1;
KW kringle domain; angiotensin; plasminogen receptor;
KW angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
KW inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
KW obesity; gene therapy; angiostatin-binding domain; Big-3.
JS Homo sapiens.
CX WO9966038-A1.
PX 23-DEC-1999.
PF 11-JUN-1999; 98WO-EP004109.
XX 15-JUN-1998; 98SE-00002130.
PR 15-JUN-1998; 98US-0089266P.
PR 17-DEC-1998; 98SE-00004372.
PR 29-DEC-1998; 98US-0114386P.
XX (PHAA) PHARMACIA & UPJOHN AB.
XX Holmgren L, Troyanovsky B;
XX WPI; 2000-106099/09.
PT Novel human protein useful for treating angiogenesis associated diseases
PT or disorders.
XX Claim 6; Page 49; 58pp; English.
XX The present sequence represents the angiotensin-binding domain,
CC designated Big-3, of a human angiogenesis-associated protein. The protein
CC binds an N-terminal fragment of plasminogen. The ABP-1 protein is
CC designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or
CC kringle 5 (K5) of plasminogen. These four kringle domains comprise
CC angiotensin. The ABP-1 protein acts as a receptor for plasminogen. A
CC polymorphic variant of ABP-1 is also described, in AAV54053. ABP-1 can be
CC used to manufacture medicaments for treating angiogenesis-related
CC diseases or disorders, such as tumor conditions, diabetes, rheumatoid
CC arthritis, and even some inflammatory diseases such as psoriasis, chronic
CC inflammation of the intestine, asthma, etc. . The protein may also be able
CC to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene
CC therapy techniques

XX Sequence 143 AA;
SQ

Query Match 21.2%; Score 143; DB 3; Length 143;
Best Local Similarity 100.0%; Pred. No. 3.7e-121;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 ESNKTAAVAPISVPAPVAAATAAATATATATTTTWWAAAPVAVAAAPAAAAAASPA 521
DB 1 ESNKTAAVAPISVPAPVAAATAAATATATATTTTWWAAAPVAVAAAPAAAAAASPA 60
QY 522 TAAATAAAVSPAAAGQIPAAASVASAAAVAPSAASAAAQVAPAPVPAPALVVPAP 581
DB 61 TAAATAAAVSPAAAGQIPAAASVASAAAVAPSAASAAAQVAPAPVPAPALVVPAP 120
QY 582 AAAQASAPAQTOAPTSAPAVAPT 604
DB 121 AAAQASAPAQTOAPTSAPAVAPT 143

RESULT 7
AAM16746
ID AAM16746 standard; protein; 208 AA.
XX AAM16746;
AC AAM16746;
DT 12-OCT-2001 (first entry)
CX Peptide #3180 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
OS Homo sapiens.
XX WO200157278-A2.
PX 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US000670.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00808408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 21572; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
CC (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 208 AA;
SQ

```
Query Match      8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred.No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPRAQPSSASYPVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 55
Db 154 MPRAQPSSASYPVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 208

RESULT 8
ABB35732
ID ABB35732 standard; peptide; 208 AA.
XX AC ABB35732;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #3238 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-483447/52.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human foetal liver.
XX PS Claim 27; SEQ ID NO 28367; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human foetal liver. The
XX CC present sequence is a peptide encoded by a single exon nucleic acid probe
XX CC of the invention. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 208 AA;

Query Match      8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred.No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPRAQPSSASYPVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 55
Db 154 MPRAQPSSASYPVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 208

RESULT 9
AAM29234
ID AAM29234 standard; protein; 208 AA.
XX AC AAM29234;
```

```
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #3271 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488897/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 27; SEQ ID NO 29503; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see A131315-A157546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders
XX SQ Sequence 208 AA;

Query Match      8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred.No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPRAQPSSASYPVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 55
Db 154 MPRAQPSSASYPVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 208

RESULT 10
ABB30567
ID ABB30567 standard; peptide; 208 AA.
XX AC ABB30567;
XX DT 01-FEB-2002 (first entry)
XX DE Peptide #3218 encoded by breast cell single exon nucleic acid probe.
XX KW Human; microarray; single exon probe; gene expression; breast; disease;
XX KW cancer.
XX OS Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000662.
```


XX Penn SG, Hanzel DX, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX Example 4; SEQ ID NO 29229; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX Sequence 208 AA;
SQ
Query Match 8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
DY 1 MRAQPSASQVPADPFAIVSRAQQWVILSDENLNRLQELGGCYEKVARLQK 55
DB 154 MRAQPSASQVPADPFAIVSRAQQWVILSDENLNRLQELGGCYEKVARLQK 208
RESULT 13
AAM56541
ID AAM56541 standard; protein; 208 AA.
AC AAM56541;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 28646.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0234359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DX, Chen W, Rank DR;
XX
WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
XX Example 4; SEQ ID NO 28646; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC

CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention
XX
SQ Sequence 208 AA;
Query Match 8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MRAQPSASQVPADPFAIVSRAQQWVILSDENLNRLQELGGCYEKVARLQK 55
DB 154 MRAQPSASQVPADPFAIVSRAQQWVILSDENLNRLQELGGCYEKVARLQK 208
RESULT 14
ABG50588
ID ABG50588 standard; peptide; 208 AA.
XX
AC ABG50588;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID NO 29236.
XX
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0234359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488998/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 29236; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47148-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 208 AA;
Query Match 8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
XX

Job time : 63 secs

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MPRAQSSASYQVPADPPAIVSRAQQMVEILSDENRNLRQELGCGYKVARLQK 55
154 MPRAQSSASYQVPADPPAIVSRAQQMVEILSDENRNLRQELGCGYKVARLQK 208

RESULT 15

AM04463
D AA004463 standard; protein; 208 AA.

X C

X C

X T

X X

X E

X X

X W

X M

X X

X S

X X

X N

X D

X X

X F

X R

X R

X R

X R

X R

X R

X R

X R

X R

X R

X A

X I

X X

X R

X X

X T

X S

X X

X C

X C

X C

09-OCT-2001 (first entry)
Peptide #3145 encoded by probe for measuring breast gene expression.

Probe; human; breast disease; breast cancer; development disorder;
inflammatory disease; proliferative breast disease; non-carcinoma tumour.

Homo sapiens.

WO200157270-A2.

09-AUG-2001.

29-JAN-2001; 2001WO-US000661.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-476286/51.

Novel single exon nucleic acid probe used to measuring gene expression in a human breast.

Claim 27; SEQ ID NO 13203; 322pp; English.

The present invention relates to novel single exon nucleic acid probes (see AA10010-AA110067). The present sequence is a peptide encoded by one such probe. The probes are useful for measuring human gene expression in a human breast sample, where the probe hybridises at high stringency to a nucleic acid expressed in the human breast. The probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human breast, particularly those diseases with polygenic aetiology. The diseases include: breast cancer; disorders of development, inflammatory diseases of the breast, fibrocystic changes, proliferative breast disease and non-carcinoma tumours. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 208 AA;

Query Match 8.1%; Score 55; DB 4; Length 208;

Best Local Similarity 100.0%; Pred. NO. 3.4e-41;

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MPRAQSSASYQVPADPPAIVSRAQQMVEILSDENRNLRQELGCGYKVARLQK 55
154 MPRAQSSASYQVPADPPAIVSRAQQMVEILSDENRNLRQELGCGYKVARLQK 208

Search completed: June 16, 2004, 19:25:42